

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```
>US-09-427-873-2
Title:
Description: (1-101) from US09427873.pap
Perfect Score: 101
Sequence: 1 LKGFQTCYNSAIGSVLTS.....STKINLDDHIANDIGTLKYE 101
```

Scoring table: TABLE unitprotable

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%

Database: pir64

```

1:pir1 2:pir2 3:pir3 4:pir4

```

Statistics: Mean 3.177; Variance 0.396; scale 8.025

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Rank	Score	Query		Length	DB	ID	Description	Pred. No.
		Match	Length					
1	101	100.0		101	2	JC5631	cyanovirin-N - Nostoc	8.77e-301
2	8	7.9		678	1	A03050	dorsal protein - fru1	1.85e-03
3	7	6.9		259	2	T07744	multicatalytic endo-	2.34e-01
4	7	6.9		261	2	A70666	probable modA protein	2.34e-01
5	7	6.9		291	1	B53419	biphenyl-2,3-diol 1,2	2.34e-01
6	7	6.9		350	2	S61646	dolichyl-diphosphooli	2.34e-01
7	7	6.9		402	2	E71912	polynucleotide adenyl	2.34e-01
8	7	6.9		440	2	B71858	adenylosuccinate lyas	2.34e-01
9	7	6.9		606	1	S59634	endo-1,4-beta-xylanas	2.34e-01
10	7	6.9		879	2	H64888	membrane protein ydB	2.34e-01
11	7	6.9		1113	2	T14260	period protein Per3 -	2.34e-01
12	7	6.9		1115	2	T13955	period protein Per3 -	2.34e-01
13	7	6.9		1200	1	SNFSO	ice nucleation protein	2.34e-01
14	6	5.9		84	2	S36691	hypothetical protein	1.85e+01
15	6	5.9		106	1	R6BY44	ribosomal protein L36	1.85e+01
16	6	5.9		107	2	A44355	CD44 glycoprotein M1	1.85e+01
17	6	5.9		109	2	H72653	hypothetical protein	1.85e+01
18	6	5.9		116	2	S63114	ribosomal protein L36	1.85e+01
19	6	5.9		122	2	S69910	Ig V-D-J region (KR)	1.85e+01
20	6	5.9		142	2	B44355	CD44 glycoprotein M2	1.85e+01
21	6	5.9		152	2	T04811	STG1 protein homolog	1.85e+01
22	6	5.9		161	2	L38108	platelet-derived grow	1.85e+01
23	6	5.9		175	2	A46388	Her-1 protein - Caeno	1.85e+01

97	6	5.9	446	2	B69307	phosphomannomutase (p	1.85e+01	170	5	5.0	14	2	PH1594	Ig H chain V-D-J regi	7.87e+02
98	6	5.9	448	2	B69745	phosphoglucosylase (g	1.85e+01	171	5	5.0	19	2	B53145	high conductance calc	7.87e+02
99	6	5.9	457	2	T02140	hypothetical protein	1.85e+01	172	5	5.0	37	2	C60580	growth hormone-relate	7.87e+02
100	6	5.9	485	2	H72424	hypothetical protein	1.85e+01	173	5	5.0	46	2	A41814	somatotropin receptor	7.87e+02
101	6	5.9	499	2	A38891	cytochrome-c oxidase	1.85e+01	174	5	5.0	46	2	S66591	hypothetical protein	7.87e+02
102	6	5.9	508	2	A48069	protein kinase MKK1 (1.85e+01	175	5	5.0	46	2	S66584	hypothetical protein	7.87e+02
103	6	5.9	522	2	I56242	lymphoid cell activat	1.85e+01	176	5	5.0	53	2	S61953	NCE1 protein - yeast	7.87e+02
104	6	5.9	520	2	S60483	GDS1 protein - yeast	1.85e+01	177	5	5.0	60	2	G70235	conserved hypothetica	7.87e+02
105	6	5.9	536	2	S06548	finger protein-binding	1.85e+01	178	5	5.0	62	2	H70081	hypothetical protein	7.87e+02
106	6	5.9	542	1	QREBOA	oligopeptide-binding	1.85e+01	179	5	5.0	63	2	G52226	cyll-S protein - Ente	7.87e+02
107	6	5.9	543	2	S58095	hypothetical protein	1.85e+01	180	5	5.0	68	2	G02089	cyll-L protein - Ente	7.87e+02
108	6	5.9	551	2	F70684	hypothetical protein	1.85e+01	181	5	5.0	68	2	A55226	bacteriocin homolog y	7.87e+02
109	6	5.9	558	2	S72592	hypothetical protein	1.85e+01	182	5	5.0	70	2	T12789	hypothetical protein	7.87e+02
110	6	5.9	561	2	T10046	hypothetical protein	1.85e+01	183	5	5.0	72	2	E07063	hypothetical protein	7.87e+02
111	6	5.9	565	2	F70192	hypothetical protein	1.85e+01	184	5	5.0	73	2	T07335	hypothetical protein	7.87e+02
112	6	5.9	565	1	HMTVGM	hemagglutinin precurs	1.85e+01	185	5	5.0	75	2	T14933	hypothetical protein	7.87e+02
113	6	5.9	566	1	HM1VT1	hemagglutinin precurs	1.85e+01	186	5	5.0	76	2	A54252	omega-agatoxin III, 8	7.87e+02
114	6	5.9	566	2	G33157	hemagglutinin precurs	1.85e+01	187	5	5.0	76	1	JC4128	CAMP-dependent protei	7.87e+02
115	6	5.9	566	1	HM1VT2	hemagglutinin precurs	1.85e+01	188	5	5.0	83	2	T16435	hypothetical protein	7.87e+02
116	6	5.9	568	2	C72129	probable omp (leader	1.85e+01	189	5	5.0	86	2	T16437	hypothetical protein	7.87e+02
117	6	5.9	569	2	T00851	hypothetical protein	1.85e+01	190	5	5.0	87	2	H31482	hypothetical protein	7.87e+02
118	6	5.9	577	2	S72227	finger protein sob -	1.85e+01	191	5	5.0	88	2	T07013	glycine-rich protein	7.87e+02
119	6	5.9	619	2	S30780	hexose metabolism-rel	1.85e+01	192	5	5.0	100	2	D72568	hypothetical protein	7.87e+02
120	6	5.9	638	1	C69126	ferrous iron transpor	1.85e+01	193	5	5.0	103	2	C71189	hypothetical protein	7.87e+02
121	6	5.9	674	2	I35476	growth potentiating f	1.85e+01	194	5	5.0	106	2	H72547	probable periplasmic	7.87e+02
122	6	5.9	688	2	S46774	multifunctional amino	1.85e+01	195	5	5.0	107	4	S57388	ubiquinol--cytochrome	7.87e+02
123	6	5.9	689	2	S17875	polynucleotide adenyl	1.85e+01	196	5	5.0	110	2	S58802	ubiquinol--cytochrome	7.87e+02
124	6	5.9	691	2	JF0150	acetylcholinesterase	1.85e+01	197	5	5.0	111	2	S58804	ubiquinol--cytochrome	7.87e+02
125	6	5.9	712	1	VCLJ54	env polypeptide precu	1.85e+01	198	5	5.0	111	2	A50666	hypothetical protein	7.87e+02
126	6	5.9	713	1	ALBSXR	cyclomaltoextrin glu	1.85e+01	199	5	5.0	111	2	F69788	hypothetical protein	7.87e+02
127	6	5.9	735	2	T12986	hypothetical protein	1.85e+01	200	5	5.0	113	2	I38216	protein-serine/threon	7.87e+02
128	6	5.9	739	2	S18644	polynucleotide adenyl	1.85e+01	201	5	5.0	114	2	G70702	hypothetical protein	7.87e+02
129	6	5.9	740	2	S17925	polynucleotide adenyl	1.85e+01	202	5	5.0	115	2	E71231	hypothetical protein	7.87e+02
130	6	5.9	754	2	S04381	beta-glucosidase (EC	1.85e+01	203	5	5.0	118	2	C71235	hypothetical protein	7.87e+02
131	6	5.9	765	2	H64606	N-methylhydantoinase	1.85e+01	204	5	5.0	122	2	A50466	hypothetical protein	7.87e+02
132	6	5.9	765	2	F71907	probable hydantoin ut	1.85e+01	205	5	5.0	122	2	F72583	microfibril-associate	7.87e+02
133	6	5.9	776	2	S44784	C30c11.4 protein - Ca	1.85e+01	206	5	5.0	127	2	PC2036	hypothetical protein	7.87e+02
134	6	5.9	779	2	A59108	beta-amylase (EC 3.2.	1.85e+01	207	5	5.0	129	2	E72773	hypothetical protein	7.87e+02
135	6	5.9	780	2	T00366	hypothetical protein	1.85e+01	208	5	5.0	132	2	E72491	hypothetical protein	7.87e+02
136	6	5.9	809	2	S67153	probable membrane pro	1.85e+01	209	5	5.0	135	2	A75139	hypothetical protein	7.87e+02
137	6	5.9	810	1	S50889	PRK1 protein - yeast	1.85e+01	210	5	5.0	138	2	D64208	hypothetical protein	7.87e+02
138	6	5.9	838	2	T05410	hypothetical protein	1.85e+01	211	5	5.0	138	2	D48188	phospholipase A2 (EC	7.87e+02
139	6	5.9	840	2	S48975	hypothetical protein	1.85e+01	212	5	5.0	140	2	C72241	interleukin-5 recepto	7.87e+02
140	6	5.9	843	2	A47132	major vault protein a	1.85e+01	213	5	5.0	140	2	C72241	conserved hypothetica	7.87e+02
141	6	5.9	859	2	F71638	env polypeptide precu	1.85e+01	214	5	5.0	143	2	H64000	hypothetical protein	7.87e+02
142	6	5.9	867	2	T15680	DNA polymerase I (pol	1.85e+01	215	5	5.0	144	1	A2P58F	azurin - pseudomonas	7.87e+02
143	6	5.9	874	2	H64228	DNA polymerase III al	1.85e+01	216	5	5.0	145	2	B35199	heat shock 16k protei	7.87e+02
144	6	5.9	914	2	T08952	hypothetical protein	1.85e+01	217	5	5.0	145	2	B24289	heat shock protein 16	7.87e+02
145	6	5.9	960	1	Q1X12	hypothetical protein	1.85e+01	218	5	5.0	147	1	MCBY	calmodulin - yeast (S	7.87e+02
146	6	5.9	966	2	S28853	phosphoenolpyruvate c	1.85e+01	219	5	5.0	148	2	G72529	hypothetical protein	7.87e+02
147	6	5.9	967	2	S25081	phosphoenolpyruvate c	1.85e+01	220	5	5.0	148	2	T14390	anther-specific prote	7.87e+02
148	6	5.9	967	2	T15680	hypothetical protein	1.85e+01	221	5	5.0	150	1	MCBB	calmodulin - rabbit	7.87e+02
149	6	5.9	1007	2	PN0156	glutamate receptor ch	1.85e+01	222	5	5.0	151	2	PQ0271	hypothetical protein	7.87e+02
150	6	5.9	1008	2	S28858	glutamate receptor de	1.85e+01	223	5	5.0	156	2	G72755	hypothetical protein	7.87e+02
151	6	5.9	1019	1	A56318	enteropeptidase (EC 3	1.85e+01	224	5	5.0	156	2	JQ1377	18k membrane protein	7.87e+02
152	6	5.9	1109	2	G75134	hypothetical protein	1.85e+01	225	5	5.0	157	2	A61017	phosphoryruvate hydra	7.87e+02
153	6	5.9	1120	2	H71664	transcription-repair	1.85e+01	226	5	5.0	157	2	S49448	oleosin - rape	7.87e+02
154	6	5.9	1192	2	T13424	hypothetical protein	1.85e+01	227	5	5.0	158	2	T16438	hypothetical protein	7.87e+02
155	6	5.9	1196	2	A29130	beta-amylase (EC 3.2.	1.85e+01	228	5	5.0	164	2	T01726	hypothetical protein	7.87e+02
156	6	5.9	1201	2	A57369	anillin - fruit fly (1.85e+01	229	5	5.0	166	2	C72734	hypothetical protein	7.87e+02
157	6	5.9	1266	2	I59314	isoleucine--trNA liga	1.85e+01	230	5	5.0	166	2	D75173	hypothetical protein	7.87e+02
158	6	5.9	1313	1	GV0VPT	M polyprotein - Punta	1.85e+01	231	5	5.0	168	2	A58883	probable transcriptio	7.87e+02
159	6	5.9	1387	2	JC5502	G-protein signaling r	1.85e+01	232	5	5.0	168	1	B69193	hypothetical protein	7.87e+02
160	6	5.9	1475	2	A44765	alpha-amylase (EC 3.2	1.85e+01	233	5	5.0	169	2	T08974	hypothetical protein	7.87e+02
161	6	5.9	1503	2	T01098	chloroplast outer env	1.85e+01	234	5	5.0	170	2	D70368	hypothetical protein	7.87e+02
162	6	5.9	1505	2	JC4851	hypoxia-inducible fac	1.85e+01	235	5	5.0	173	1	D2Y25X	development-specific	7.87e+02
163	6	5.9	1676	2	E71410	probable centromere p	1.85e+01	236	5	5.0	174	2	G70877	probable reex protein	7.87e+02
164	6	5.9	1732	2	E71442	hypothetical protein	1.85e+01	237	5	5.0	175	2	B44102	di-N-acetylchitinobiase	7.87e+02
165	6	5.9	1890	2	T04556	hypothetical protein	1.85e+01	238	5	5.0	175	2	B70783	hypothetical protein	7.87e+02
166	6	5.9	1992	2	A47237	myosin heavy chain fo	1.85e+01	239	5	5.0	175	1	D2Y21X	gene 1 protein - Myxo	7.87e+02
167	6	5.9	3707	2	S18252	heparan sulfate prote	1.85e+01	240	5	5.0	176	2	F71540	hypothetical protein	7.87e+02
168	6	5.9	3759	2	A35085	trithorax protein - f	1.85e+01	241	5	5.0	177	2	A49465	coatmer zeta chain -	7.87e+02
169	6	5.9	3828	2	T13857	trithorax protein - f	1.85e+01	242	5	5.0	178	2	S71891	immune factor 1, type	7.87e+02

243	5	5.0	180	1	GGGP	glucagon precursor -	7.87e+02	316	5	5.0	257	2	I64181	probable amino acid-b	7.87e+02
244	5	5.0	181	2	E72014	peptidyl tRNA hydrola	7.87e+02	317	5	5.0	257	2	B71892	flagellar motor prote	7.87e+02
245	5	5.0	182	2	T13821	NADH dehydrogenase su	7.87e+02	318	5	5.0	259	2	D71412	probable imidazolegly	7.87e+02
246	5	5.0	183	2	D70339	conserved hypothetical	7.87e+02	319	5	5.0	260	2	B75181	proteasome, chain alp	7.87e+02
247	5	5.0	183	2	S33422	hypothetical protein	7.87e+02	320	5	5.0	263	2	G71114	probable ORF2 protein	7.87e+02
248	5	5.0	188	2	F72601	hypothetical protein	7.87e+02	321	5	5.0	263	2	E75046	hypothetical protein	7.87e+02
249	5	5.0	188	2	A39185	virulence protein pag	7.87e+02	322	5	5.0	263	1	SNRTC2	multicatalytic endope	7.87e+02
250	5	5.0	189	1	VCTV	coat protein - turnip	7.87e+02	323	5	5.0	266	2	A34466	calpain (EC 3.4.22.17	7.87e+02
251	5	5.0	192	2	S26199	hypothetical protein	7.87e+02	324	5	5.0	266	2	E71230	hypothetical protein	7.87e+02
252	5	5.0	194	2	A72727	probable ribosomal pr	7.87e+02	325	5	5.0	266	1	C1PGL	calpain (EC 3.4.22.17	7.87e+02
253	5	5.0	194	2	B37854	para-aminobenzoate sy	7.87e+02	326	5	5.0	267	2	S44225	strf protein - Strept	7.87e+02
254	5	5.0	195	2	S77753	peptide transport sys	7.87e+02	327	5	5.0	268	2	F64024	hypothetical protein	7.87e+02
255	5	5.0	196	2	JC2319	replication protein -	7.87e+02	328	5	5.0	269	1	JC1445	multicatalytic endope	7.87e+02
256	5	5.0	197	2	B64387	imidazoleglycerol-pho	7.87e+02	329	5	5.0	269	2	D64668	hypothetical protein	7.87e+02
257	5	5.0	198	2	J00864	hypothetical 21.9K pr	7.87e+02	330	5	5.0	270	2	F72259	hypothetical protein	7.87e+02
258	5	5.0	198	2	T06261	probable imidazolegly	7.87e+02	331	5	5.0	273	2	T01985	zinc-finger protein	7.87e+02
259	5	5.0	200	2	A41740	H+-transporting ATP s	7.87e+02	332	5	5.0	277	2	S76356	proteinase IV (EC 3.4	7.87e+02
260	5	5.0	202	2	C71377	hypothetical protein	7.87e+02	333	5	5.0	280	2	E72745	hypothetical protein	7.87e+02
261	5	5.0	204	2	B27333	probable membrane-ass	7.87e+02	334	5	5.0	282	2	A41025	asparaginase II (7.87e+02
262	5	5.0	204	2	S22639	alpha-amylase/subtili	7.87e+02	335	5	5.0	282	2	H70006	N-acetylmuramoyl-L-al	7.87e+02
263	5	5.0	206	2	T14880	hypothetical protein	7.87e+02	336	5	5.0	282	2	T06630	growth factor like pr	7.87e+02
264	5	5.0	207	2	E69184	hypothetical protein	7.87e+02	337	5	5.0	283	1	B49769	gap junction protein	7.87e+02
265	5	5.0	209	2	S55887	CCHH finger protein 7	7.87e+02	338	5	5.0	284	2	A71210	hypothetical protein	7.87e+02
266	5	5.0	211	2	S72580	hypothetical protein	7.87e+02	339	5	5.0	285	2	JC4315	steroidogenic acute r	7.87e+02
267	5	5.0	213	2	A71049	hypothetical protein	7.87e+02	340	5	5.0	285	2	I38248	transcription factor	7.87e+02
268	5	5.0	215	2	S70837	pilo protein - Neisse	7.87e+02	341	5	5.0	286	2	C72098	SUA5 related protein	7.87e+02
269	5	5.0	216	2	H72291	hypothetical protein	7.87e+02	342	5	5.0	288	2	T05954	transcription factor	7.87e+02
270	5	5.0	217	1	H71208	hypothetical protein	7.87e+02	343	5	5.0	289	2	E72625	probable acetylglutam	7.87e+02
271	5	5.0	218	2	F69972	probable membrane pro	7.87e+02	344	5	5.0	289	1	S55649	deoxyuridine triphosp	7.87e+02
272	5	5.0	218	2	T15321	hypothetical protein	7.87e+02	345	5	5.0	291	2	F71416	hypothetical protein	7.87e+02
273	5	5.0	220	2	I40539	vsr protein - Pseudo	7.87e+02	346	5	5.0	292	2	A46352	ORF5 protein - Chlore	7.87e+02
274	5	5.0	221	2	S69682	hypothetical protein	7.87e+02	347	5	5.0	292	2	S11870	peroxidase (EC 1.11.1	7.87e+02
275	5	5.0	222	2	D69014	conserved hypothetical	7.87e+02	348	5	5.0	293	2	B71012	hypothetical protein	7.87e+02
276	5	5.0	223	2	S48832	superoxide dismutase	7.87e+02	349	5	5.0	294	2	S23561	HALL protein - yeast	7.87e+02
277	5	5.0	223	1	YUHP	serum amyloid P-compo	7.87e+02	350	5	5.0	294	2	S71642	dolichyl phosphoryl m	7.87e+02
278	5	5.0	223	2	T07098	probable catechol oxi	7.87e+02	351	5	5.0	296	2	S16320	light-induced protein	7.87e+02
279	5	5.0	223	1	D69305	conserved hypothetical	7.87e+02	352	5	5.0	296	2	S62624	biliverdin reductase	7.87e+02
280	5	5.0	224	2	F71329	probable phosphoglyco	7.87e+02	353	5	5.0	299	1	J00613	3-hydroxyisobutyrate	7.87e+02
281	5	5.0	225	2	B72587	hypothetical protein	7.87e+02	354	5	5.0	302	2	S71890	immune-factor 1, type	7.87e+02
282	5	5.0	226	2	D70540	probable dethiobiotin	7.87e+02	355	5	5.0	302	1	R3SPS5	ribosomal protein CS-	7.87e+02
283	5	5.0	226	2	A70565	probable cutinase pre	7.87e+02	356	5	5.0	304	2	A49185	interphotoreceptor re	7.87e+02
284	5	5.0	227	2	T00001	pepf protein - Acidia	7.87e+02	357	5	5.0	305	2	I38144	phosphotyrosine picke	7.87e+02
285	5	5.0	228	1	M5E170	histidine-c oxidase	7.87e+02	358	5	5.0	305	2	C39240	finger protein mfg3	7.87e+02
286	5	5.0	228	1	MEBQ7	histidine permease pr	7.87e+02	359	5	5.0	305	2	A75211	plant- type 1-asparag	7.87e+02
287	5	5.0	229	2	C43431	phosphorylase kinase	7.87e+02	360	5	5.0	305	2	A71247	probable L-asparagina	7.87e+02
288	5	5.0	230	2	S27996	outer membrane protei	7.87e+02	361	5	5.0	306	2	S22523	asparaginase (EC 3.5	7.87e+02
289	5	5.0	231	2	A31572	LuxF protein - Photob	7.87e+02	362	5	5.0	307	2	S26215	ornithine carbamoylir	7.87e+02
290	5	5.0	233	2	G75075	hypothetical protein	7.87e+02	363	5	5.0	308	2	G71454	hypothetical protein	7.87e+02
291	5	5.0	235	2	B42337	parotid secretory pro	7.87e+02	364	5	5.0	308	2	E72627	hypothetical protein	7.87e+02
292	5	5.0	236	2	H69619	hypothetical protein	7.87e+02	365	5	5.0	310	2	E72667	hypothetical protein	7.87e+02
293	5	5.0	236	2	S60390	probable membrane pro	7.87e+02	366	5	5.0	311	2	S66011	transcription regulat	7.87e+02
294	5	5.0	238	2	S24267	cooB protein - Escher	7.87e+02	367	5	5.0	312	2	T08985	hypothetical protein	7.87e+02
295	5	5.0	239	2	H72667	hypothetical protein	7.87e+02	368	5	5.0	313	2	S66312	G-box binding factor	7.87e+02
296	5	5.0	239	2	S09277	sensory rhodopsin I -	7.87e+02	369	5	5.0	314	1	PWBYG	H+-transporting ATP s	7.87e+02
297	5	5.0	241	2	B70407	alcohol dehydrogenase	7.87e+02	370	5	5.0	315	2	S20883	G-box-binding factor	7.87e+02
298	5	5.0	241	2	F73747	glycerophosphoryl die	7.87e+02	371	5	5.0	315	2	C64092	hypothetical protein	7.87e+02
299	5	5.0	241	2	I51211	cytoskeletal protein	7.87e+02	372	5	5.0	316	1	PWYCG	H+-transporting ATP s	7.87e+02
300	5	5.0	244	2	H70591	hypothetical protein	7.87e+02	373	5	5.0	319	2	A70340	conserved hypothetical	7.87e+02
301	5	5.0	244	2	A49202	cysteine-rich secreto	7.87e+02	374	5	5.0	319	2	H72391	K+ channel, beta subu	7.87e+02
302	5	5.0	248	2	B39534	MADS box protein AGL2	7.87e+02	375	5	5.0	319	2	T15137	hypothetical protein	7.87e+02
303	5	5.0	250	2	G70758	hypothetical protein	7.87e+02	376	5	5.0	320	2	D69203	methenyltetrahydromet	7.87e+02
304	5	5.0	251	2	S39202	cellulase (EC 3.2.1.4	7.87e+02	377	5	5.0	321	2	T13539	hypothetical protein	7.87e+02
305	5	5.0	251	2	H71638	hypothetical protein	7.87e+02	378	5	5.0	325	2	S18575	syrm protein - Rhizob	7.87e+02
306	5	5.0	252	1	B45539	matrix protein M1 - i	7.87e+02	379	5	5.0	325	2	S54873	UDP-N-acetylmuramoyla	7.87e+02
307	5	5.0	252	1	MFIV61	matrix protein M1 - i	7.87e+02	380	5	5.0	326	2	A44505	syrm protein - Rhizob	7.87e+02
308	5	5.0	252	1	MFIV1K	matrix protein M1 - i	7.87e+02	381	5	5.0	327	2	D72503	probable modification	7.87e+02
309	5	5.0	252	2	S61178	hypothetical protein	7.87e+02	382	5	5.0	328	2	S32369	gamma-SNAP protein -	7.87e+02
310	5	5.0	252	1	MFIVWS	matrix protein M1 - i	7.87e+02	383	5	5.0	330	2	T07394	probable potassium ch	7.87e+02
311	5	5.0	252	1	MFIV1M	matrix protein M1 - i	7.87e+02	384	5	5.0	330	2	A64526	hypothetical protein	7.87e+02
312	5	5.0	252	1	MFIV1F	matrix protein M1 - i	7.87e+02	385	5	5.0	330	2	E64585	viRb1 homolog - Heli	7.87e+02
313	5	5.0	253	2	A31392	chlorophyll a/b-bindi	7.87e+02	386	5	5.0	331	2	S45584	transcription activat	7.87e+02
314	5	5.0	255	2	T11761	pod storage protein -	7.87e+02	387	5	5.0	336	2	H71859	glucokinase - Helicob	7.87e+02
315	5	5.0	256	2	S34850	hypothetical protein	7.87e+02	388	5	5.0	337	2	S77630	hypothetical protein	7.87e+02

389	5.0	339	2	H71361	conserved hypothetical	7.87e+02	462	5	5.0	386	2	T16099	hypothetical protein	7.87e+02
390	5.0	339	2	D75069	hypothetical protein	7.87e+02	463	5	5.0	386	1	S33852	hydrogenase (EC 1.18.	7.87e+02
391	5.0	339	2	D70942	probable pfkB protein	7.87e+02	464	5	5.0	387	2	T07591	polygalacturonase (EC	7.87e+02
392	5.0	340	2	A72600	probable dehydrogenas	7.87e+02	465	5	5.0	387	2	T04320	polygalacturonase (EC	7.87e+02
393	5.0	341	2	D72622	hypothetical protein	7.87e+02	466	5	5.0	388	1	MMVZM1	major envelope antigen	7.87e+02
394	5.0	344	2	H71286	hypothetical protein	7.87e+02	467	5	5.0	388	1	WMVZU2	probable high-affinit	7.87e+02
395	5.0	346	2	T14020	NADH dehydrogenase su	7.87e+02	468	5	5.0	390	2	D72688	branched-chain amino	7.87e+02
396	5.0	346	2	T14049	NADH dehydrogenase su	7.87e+02	469	5	5.0	390	2	B69370	hypothetical protein	7.87e+02
397	5.0	346	2	T14133	NADH dehydrogenase su	7.87e+02	470	5	5.0	392	2	S72984	undecaprenol-phosphat	7.87e+02
398	5.0	346	2	T14038	NADH dehydrogenase su	7.87e+02	471	5	5.0	392	2	S39657	hypothetical protein	7.87e+02
399	5.0	346	2	T13860	NADH dehydrogenase su	7.87e+02	472	5	5.0	392	2	T01849	polygalacturonase pre	7.87e+02
400	5.0	346	2	T14055	NADH dehydrogenase su	7.87e+02	473	5	5.0	392	2	S37806	MF1 protein - yeast	7.87e+02
401	5.0	346	2	T14033	NADH dehydrogenase su	7.87e+02	474	5	5.0	392	2	S32405	sedoheptulose-bisphos	7.87e+02
402	5.0	346	2	T14032	NADH dehydrogenase su	7.87e+02	475	5	5.0	393	2	S51838	bexd protein - haemo	7.87e+02
403	5.0	346	2	T14056	NADH dehydrogenase su	7.87e+02	476	5	5.0	394	1	BWHIXD	probable glutamate/ a	7.87e+02
404	5.0	346	2	T13973	NADH dehydrogenase su	7.87e+02	477	5	5.0	396	2	G71309	hypothetical protein	7.87e+02
405	5.0	346	2	T14100	NADH dehydrogenase su	7.87e+02	478	5	5.0	396	2	S38924	DNA-directed RNA poly	7.87e+02
406	5.0	346	2	T13835	NADH dehydrogenase su	7.87e+02	479	5	5.0	397	2	C75182	DNA-directed RNA poly	7.87e+02
407	5.0	346	2	T13920	NADH dehydrogenase su	7.87e+02	480	5	5.0	397	2	S03577	hypothetical protein	7.87e+02
408	5.0	346	2	T13834	NADH dehydrogenase su	7.87e+02	481	5	5.0	398	2	H64534	paired-type homeodoma	7.87e+02
409	5.0	347	2	S57183	hypothetical protein	7.87e+02	482	5	5.0	399	2	JC6522	LIM domain-containing	7.87e+02
410	5.0	347	2	T14204	NADH dehydrogenase su	7.87e+02	483	5	5.0	400	2	JC6522	hypothetical protein	7.87e+02
411	5.0	347	2	T14205	NADH dehydrogenase su	7.87e+02	484	5	5.0	402	2	T12745	bicyclomycin resistan	7.87e+02
412	5.0	347	2	T14057	probable peptide tran	7.87e+02	485	5	5.0	402	2	E69783	hypothetical protein	7.87e+02
413	5.0	347	2	C72452	NADH dehydrogenase su	7.87e+02	486	5	5.0	405	2	D72426	phosphoprotein phosph	7.87e+02
414	5.0	347	2	T13966	NADH dehydrogenase su	7.87e+02	487	5	5.0	406	2	A47492	BZIP DNA-binding prot	7.87e+02
415	5.0	347	2	T14129	NADH dehydrogenase su	7.87e+02	488	5	5.0	407	2	T14909	probable monooxygenas	7.87e+02
416	5.0	347	2	A46567	tetracycline resistanc	7.87e+02	489	5	5.0	407	2	C75218	probable hydroxynitri	7.87e+02
417	5.0	348	2	T11765	NADH dehydrogenase (u	7.87e+02	490	5	5.0	407	2	G71414	hypothetical protein	7.87e+02
418	5.0	348	2	T11288	NADH dehydrogenase (u	7.87e+02	491	5	5.0	408	2	B71272	probable lpqk protein	7.87e+02
419	5.0	349	2	T15422	hypothetical protein	7.87e+02	492	5	5.0	409	2	F70633	probable acetyl-CoA c	7.87e+02
420	5.0	350	1	VGBE63	glycoprotein gp63 - s	7.87e+02	493	5	5.0	411	2	T07261	oxidoreductase homolo	7.87e+02
421	5.0	351	2	A45377	transcription factor	7.87e+02	494	5	5.0	411	2	T01388	carboxypeptidase III	7.87e+02
422	5.0	351	2	T12728	structural protein -	7.87e+02	495	5	5.0	411	2	A35275	probable Na+/H+ anti	7.87e+02
423	5.0	351	2	I46033	NIPP-1, nuclear inhib	7.87e+02	496	5	5.0	412	2	A72473	erythromycin monooxyg	7.87e+02
424	5.0	351	2	G62783	UDPglucose 4-epimeras	7.87e+02	497	5	5.0	412	1	B40634	hypothetical protein	7.87e+02
425	5.0	351	1	SAVZVV	surface antigen precu	7.87e+02	498	5	5.0	412	2	D71972	probable carbohydrate	7.87e+02
426	5.0	352	2	C70578	hypothetical protein	7.87e+02	499	5	5.0	414	2	B71345	penicillin-binding pr	7.87e+02
427	5.0	353	2	S41958	pupal cuticle protei	7.87e+02	500	5	5.0	414	2	S75832	dynamin heavy chain -	7.87e+02
428	5.0	353	2	S69730	hypothetical protein	7.87e+02	501	5	5.0	416	2	S49584	probable membrane pro	7.87e+02
429	5.0	353	1	SAVZVC	surface antigen precu	7.87e+02	502	5	5.0	417	2	S19382	triacylglycerol lipas	7.87e+02
430	5.0	355	2	C39725	hypothetical protein	7.87e+02	503	5	5.0	418	2	H72203	hypothetical protein	7.87e+02
431	5.0	356	2	F72742	probable transport AT	7.87e+02	504	5	5.0	420	2	S21052	interleukin-5 recepto	7.87e+02
432	5.0	356	2	A75044	translation initiatio	7.87e+02	505	5	5.0	420	2	A64220	hypothetical protein	7.87e+02
433	5.0	357	2	S75954	hypothetical protein	7.87e+02	506	5	5.0	421	2	A47713	chitin deacetylase (E	7.87e+02
434	5.0	358	1	PKWKG	H+-transporting ATP s	7.87e+02	507	5	5.0	421	2	T05583	CER2 protein - Arabid	7.87e+02
435	5.0	359	3	T12259	O-diphenol O-methylr	7.87e+02	508	5	5.0	424	2	T10985	regulator protein ROM	7.87e+02
436	5.0	359	2	B64641	nickel-cobalt-cadmium	7.87e+02	509	5	5.0	424	2	C69077	thiamine biosynthesis	7.87e+02
437	5.0	363	2	T10638	hypothetical protein	7.87e+02	510	5	5.0	426	2	A70444	flagellar hook associ	7.87e+02
438	5.0	364	2	B71139	hypothetical protein	7.87e+02	511	5	5.0	426	2	S26143	rad9 protein (allele	7.87e+02
439	5.0	364	2	G72513	hypothetical protein	7.87e+02	512	5	5.0	428	2	B71278	hypothetical protein	7.87e+02
440	5.0	364	2	S36403	catechol O-methyltran	7.87e+02	513	5	5.0	428	2	S03607	hypothetical protein	7.87e+02
441	5.0	365	2	T08577	hypothetical protein	7.87e+02	514	5	5.0	429	2	T03607	serine carboxypeptida	7.87e+02
442	5.0	365	2	C37753	quinolinate synthase	7.87e+02	515	5	5.0	429	2	S41527	citrate (si)-synthase	7.87e+02
443	5.0	366	2	G71067	hypothetical protein	7.87e+02	516	5	5.0	430	2	A56209	transcription factor	7.87e+02
444	5.0	366	2	A75077	iron (iii) abc transp	7.87e+02	517	5	5.0	430	1	CMUMF	mucorpepsin (EC 3.4.2	7.87e+02
445	5.0	366	2	A71880	probable outer membra	7.87e+02	518	5	5.0	430	2	F64892	probable membrane pro	7.87e+02
446	5.0	366	2	A70852	hypothetical protein	7.87e+02	519	5	5.0	431	2	E71050	hypothetical protein	7.87e+02
447	5.0	367	2	E64579	outer membrane protei	7.87e+02	520	5	5.0	431	2	S30116	preproteins translocas	7.87e+02
448	5.0	369	2	C64635	outer membrane protei	7.87e+02	521	5	5.0	431	2	S77340	hypothetical protein	7.87e+02
449	5.0	370	2	D75084	carotenoid biosynthet	7.87e+02	522	5	5.0	433	2	F70337	hypothetical protein	7.87e+02
450	5.0	371	2	S75778	oligopeptide transpor	7.87e+02	523	5	5.0	439	2	E71268	conserved hypothetical	7.87e+02
451	5.0	371	1	S59442	DNA repair protein RA	7.87e+02	524	5	5.0	440	2	A40624	phosphopyruvate hydra	7.87e+02
452	5.0	373	1	B76273	hypothetical protein	7.87e+02	525	5	5.0	440	2	A37044	alpha-1,6-mannosyl-gl	7.87e+02
453	5.0	373	2	B39732	H+-transporting ATP s	7.87e+02	526	5	5.0	442	2	S70648	gene feebly protein -	7.87e+02
454	5.0	374	2	B34800	thin filament-associa	7.87e+02	527	5	5.0	445	2	A64092	acetyl-CoA C-acetyltr	7.87e+02
455	5.0	378	2	T10327	hypothetical protein	7.87e+02	528	5	5.0	445	2	B25076	tubulin alpha-3 chain	7.87e+02
456	5.0	380	2	S34964	rbc protein - Shigell	7.87e+02	529	5	5.0	445	2	S25554	tubulin beta chain -	7.87e+02
457	5.0	381	2	S26812	transcription factor	7.87e+02	530	5	5.0	446	2	S25256	alpha-1,6-mannosyl-gl	7.87e+02
458	5.0	381	2	H69096	bacteriochlorophyll s	7.87e+02	531	5	5.0	447	2	S52751	heat shock transcript	7.87e+02
459	5.0	381	2	T03923	probable omega-3 fatt	7.87e+02	532	5	5.0	448	2	D70179	Na+/H+ antiporter (nh	7.87e+02
460	5.0	382	2	T10289	hypothetical protein	7.87e+02	533	5	5.0	449	2	T05242	hypothetical protein	7.87e+02
461	5.0	386	2	T00589	hypothetical protein	7.87e+02	534	5	5.0	449	2			

535	5	5.0	533	5	5.0	608	inner membrane protein	7.87e+02	beta-fructofuranosida	7.87e+02
536	5	5.0	536	5	5.0	609	probable preprotein t	7.87e+02	cytochrome-c oxidase	7.87e+02
537	5	5.0	537	5	5.0	610	CP49ins protein - chi	7.87e+02	probable capsular pol	7.87e+02
538	5	5.0	538	5	5.0	611	hypothetical protein	7.87e+02	probable csp protein	7.87e+02
539	5	5.0	539	5	5.0	612	indole-3-glycerol-pho	7.87e+02	PEP12 protein - yeas	7.87e+02
540	5	5.0	540	5	5.0	613	probable glucosyltran	7.87e+02	L-glutamate transport	7.87e+02
541	5	5.0	541	5	5.0	614	acetyl-CoA decarboxyl	7.87e+02	hypothetical protein	7.87e+02
542	5	5.0	542	5	5.0	615	sensory transduction	7.87e+02	mucin JUL7 - human	7.87e+02
543	5	5.0	543	5	5.0	616	indole-3-glycerol-pho	7.87e+02	carboxylesterase (EC	7.87e+02
544	5	5.0	544	5	5.0	617	flagellar hook protel	7.87e+02	neuroglycan C precurs	7.87e+02
545	5	5.0	545	5	5.0	618	probable glutamate de	7.87e+02	sterol carrier protei	7.87e+02
546	5	5.0	546	5	5.0	619	glutamyL-tRNA reducta	7.87e+02	parathyroid hormone r	7.87e+02
547	5	5.0	547	5	5.0	620	hexokinase (EC 2.7.1.1	7.87e+02	glucose-6-phosphate 1	7.87e+02
548	5	5.0	548	5	5.0	621	chromosomal replicati	7.87e+02	hypothetical protein	7.87e+02
549	5	5.0	549	5	5.0	622	hypothetical protein	7.87e+02	protein-tyrosine kina	7.87e+02
550	5	5.0	550	5	5.0	623	acetyl-CoA C-acyltran	7.87e+02	serine C-palmitoyltra	7.87e+02
551	5	5.0	551	5	5.0	624	hydroxymethylglutaryl	7.87e+02	probable apolipoprote	7.87e+02
552	5	5.0	552	5	5.0	625	site-specific DNA-met	7.87e+02	site-specific DNA-met	7.87e+02
553	5	5.0	553	5	5.0	626	sel-12 protein - Caen	7.87e+02	polyubiquitin-like pr	7.87e+02
554	5	5.0	554	5	5.0	627	arginine catabolism p	7.87e+02	hemagglutinin precurs	7.87e+02
555	5	5.0	555	5	5.0	628	probable grp binding	7.87e+02	glucan 1,3-beta-glucos	7.87e+02
556	5	5.0	556	5	5.0	629	tubulin alpha-4 chain	7.87e+02	thin filament-associat	7.87e+02
557	5	5.0	557	5	5.0	630	1-aminocyclopropane-1	7.87e+02	bacillolysin (EC 3.4.	7.87e+02
558	5	5.0	558	5	5.0	631	yalu protein - Escher	7.87e+02	Rab geranylgeranyl tr	7.87e+02
559	5	5.0	559	5	5.0	632	conserved hypotheticala	7.87e+02	probable 3'-phosphoad	7.87e+02
560	5	5.0	560	5	5.0	633	fumarate reductase (E	7.87e+02	proline--tRNA ligase	7.87e+02
561	5	5.0	561	5	5.0	634	IGA-specific metallo	7.87e+02	X11L2 protein - human	7.87e+02
562	5	5.0	562	5	5.0	635	gene forked protein -	7.87e+02	cerebroglycan precurs	7.87e+02
563	5	5.0	563	5	5.0	636	hypothetical protein	7.87e+02	probable alpha-amylas	7.87e+02
564	5	5.0	564	5	5.0	637	hypothetical protein	7.87e+02	hypothetical protein	7.87e+02
565	5	5.0	565	5	5.0	638	hypothetical protein	7.87e+02	zyxin - human	7.87e+02
566	5	5.0	566	5	5.0	639	66K glycoprotein prec	7.87e+02	probable 3'-phosphoad	7.87e+02
567	5	5.0	567	5	5.0	640	probable 3-hydroxybut	7.87e+02	X11L2 protein - human	7.87e+02
568	5	5.0	568	5	5.0	641	hypothetical protein	7.87e+02	topoisomerase I-relat	7.87e+02
569	5	5.0	569	5	5.0	642	major plasmid transfe	7.87e+02	hypothetical protein	7.87e+02
570	5	5.0	570	5	5.0	643	subtilisin-like prote	7.87e+02	probable outer membra	7.87e+02
571	5	5.0	571	5	5.0	644	mannosyltransferase (7.87e+02	hypothetical protein	7.87e+02
572	5	5.0	572	5	5.0	645	hypothetical protein	7.87e+02	radixin - mouse	7.87e+02
573	5	5.0	573	5	5.0	646	protein kinase - mous	7.87e+02	archaeosine trna-ribo	7.87e+02
574	5	5.0	574	5	5.0	647	hypothetical protein	7.87e+02	topoisomerase I-relat	7.87e+02
575	5	5.0	575	5	5.0	648	hypothetical protein	7.87e+02	hypothetical protein	7.87e+02
576	5	5.0	576	5	5.0	649	cytochrome-c oxidase	7.87e+02	probable membrane pro	7.87e+02
577	5	5.0	577	5	5.0	650	hypothetical protein	7.87e+02	hypothetical protein	7.87e+02
578	5	5.0	578	5	5.0	651	xylokinase homolog	7.87e+02	two-component sensor	7.87e+02
579	5	5.0	579	5	5.0	652	phosphorin III - Vol	7.87e+02	hypothetical protein	7.87e+02
580	5	5.0	580	5	5.0	653	hook-associated prote	7.87e+02	probable alpha-amylas	7.87e+02
581	5	5.0	581	5	5.0	654	penicillin binding pr	7.87e+02	hypothetical protein	7.87e+02
582	5	5.0	582	5	5.0	655	X-Pro dipeptidase (EC	7.87e+02	catechol oxidase (EC	7.87e+02
583	5	5.0	583	5	5.0	656	X-Pro dipeptidase (EC	7.87e+02	cyclomaltodextrinase	7.87e+02
584	5	5.0	584	5	5.0	657	cholesteryl ester tra	7.87e+02	apolipoprotein B - cr	7.87e+02
585	5	5.0	585	5	5.0	658	6-aminohexanoate-cycl	7.87e+02	catechol oxidase (EC	7.87e+02
586	5	5.0	586	5	5.0	659	LIC-2 - rat	7.87e+02	mucin - human (fragme	7.87e+02
587	5	5.0	587	5	5.0	660	ammonium transport pr	7.87e+02	NADH dehydrogenase su	7.87e+02
588	5	5.0	588	5	5.0	661	phosphatidylserine de	7.87e+02	hypothetical protein	7.87e+02
589	5	5.0	589	5	5.0	662	hexose phosphate tran	7.87e+02	phosphoprotein phosph	7.87e+02
590	5	5.0	590	5	5.0	663	myosin heavy chain, e	7.87e+02	threonyl-tRNA synthet	7.87e+02
591	5	5.0	591	5	5.0	664	sodium/proline sympor	7.87e+02	hemagglutinin-neuram1	7.87e+02
592	5	5.0	592	5	5.0	665	zinc transporter Znt-	7.87e+02	probable RNA-directed	7.87e+02
593	5	5.0	593	5	5.0	666	heat shock transcript	7.87e+02	hypothetical protein	7.87e+02
594	5	5.0	594	5	5.0	667	agarase (EC 3.2.1.81)	7.87e+02	probable phosphoenolp	7.87e+02
595	5	5.0	595	5	5.0	668	hypothetical protein	7.87e+02	phosphoprotein phosph	7.87e+02
596	5	5.0	596	5	5.0	669	zinc transport protei	7.87e+02	threonyl-tRNA synthet	7.87e+02
597	5	5.0	597	5	5.0	670	hypothetical protein	7.87e+02	hypothetical protein	7.87e+02
598	5	5.0	598	5	5.0	671	aromatic-L-amino-acid	7.87e+02	ARP protein - Arabido	7.87e+02
599	5	5.0	599	5	5.0	672	probable membrane pro	7.87e+02	hypothetical protein	7.87e+02
600	5	5.0	600	5	5.0	673	hypothetical protein	7.87e+02	hypothetical protein	7.87e+02
601	5	5.0	601	5	5.0	674	hypothetical protein	7.87e+02	hypothetical protein	7.87e+02
602	5	5.0	602	5	5.0	675	probable minor struct	7.87e+02	hypothetical protein	7.87e+02
603	5	5.0	603	5	5.0	676	nicotinic acetylcholi	7.87e+02	hypothetical protein	7.87e+02
604	5	5.0	604	5	5.0	677	probable lipoygenase	7.87e+02	hypothetical protein	7.87e+02
605	5	5.0	605	5	5.0	678	probable DNA repair p	7.87e+02	hypothetical protein	7.87e+02
606	5	5.0	606	5	5.0	679	unspecific monooxygen	7.87e+02	hypothetical protein	7.87e+02
607	5	5.0	607	5	5.0	680	hypothetical protein	7.87e+02	Viral replicase 2 - b	7.87e+02

681	5	5.0	638	2	A33991	somatotropin receptor	7.87e+02	754	5	5.0	758	2	T10614	hypothetical protein	7.87e+02
682	5	5.0	638	2	B28176	somatotropin receptor	7.87e+02	755	5	5.0	759	2	G71279	probable exoribonucle	7.87e+02
683	5	5.0	638	2	S04530	somatotropin receptor	7.87e+02	756	5	5.0	762	2	T14815	hypothetical protein	7.87e+02
684	5	5.0	638	2	A31845	nitrous-oxide reducta	7.87e+02	757	5	5.0	763	2	A29562	PRT1 protein - yeast	7.87e+02
685	5	5.0	638	2	S33884	DNA-directed RNA poly	7.87e+02	758	5	5.0	764	2	B70979	hypothetical protein	7.87e+02
686	5	5.0	638	2	E70528	probable dxs protein	7.87e+02	759	5	5.0	766	2	G71845	probable role in oute	7.87e+02
687	5	5.0	641	2	T05497	hypothetical protein	7.87e+02	760	5	5.0	767	2	S47639	acetylcholinesterase	7.87e+02
688	5	5.0	642	2	F64866	probable transcriptio	7.87e+02	761	5	5.0	771	1	WV29J	ribonucleoside-diphos	7.87e+02
689	5	5.0	642	2	S59306	probable membrane pro	7.87e+02	762	5	5.0	771	2	B36843	ribonucleoside-diphos	7.87e+02
690	5	5.0	647	2	J02149	B west mating protein	7.87e+02	763	5	5.0	773	2	J60387	exo-alpha-sialidase (7.87e+02
691	5	5.0	648	2	T12570	NADH dehydrogenase su	7.87e+02	764	5	5.0	774	2	T14555	DNA polymerase homolo	7.87e+02
692	5	5.0	653	2	S40962	DNA polymerase alpha	7.87e+02	765	5	5.0	776	1	VPXRRH	outer layer protein V	7.87e+02
693	5	5.0	655	2	A46270	Na and Cl dependent t	7.87e+02	766	5	5.0	776	1	VPXRB3	outer layer protein V	7.87e+02
694	5	5.0	660	2	H64671	conserved hypothetica	7.87e+02	767	5	5.0	776	1	VPXRT1	outer layer protein V	7.87e+02
695	5	5.0	670	2	A25001	catalase (EC 1.11.1.6	7.87e+02	768	5	5.0	776	2	S24410	hypothetical outer ca	7.87e+02
696	5	5.0	673	2	T15551	hypothetical protein	7.87e+02	769	5	5.0	777	2	S65543	3',5'-cyclic-nucleoti	7.87e+02
697	5	5.0	674	2	T05264	probable serine/threo	7.87e+02	770	5	5.0	780	2	A48143	HF-1 regulatory eleme	7.87e+02
698	5	5.0	677	2	S33608	somatotropin-binding	7.87e+02	771	5	5.0	782	2	S04047	finger protein zfy-1	7.87e+02
699	5	5.0	677	2	C42125	trophozoite cysteine-	7.87e+02	772	5	5.0	782	2	A31491	sex-determining regio	7.87e+02
700	5	5.0	678	2	B48089	growth arrest-specifi	7.87e+02	773	5	5.0	784	2	S26638	SPR-1 protein - human	7.87e+02
701	5	5.0	679	2	I48084	gibbon ape leukemia r	7.87e+02	774	5	5.0	786	2	H69980	single-strand DNA-spe	7.87e+02
702	5	5.0	682	2	T12294	NADH dehydrogenase -	7.87e+02	775	5	5.0	789	1	A45617	6-phosphofructokinase	7.87e+02
703	5	5.0	682	2	T04846	protein kinase homolo	7.87e+02	776	5	5.0	790	2	I50178	cadherin-6B - chicken	7.87e+02
704	5	5.0	683	2	I52996	transforming growth f	7.87e+02	777	5	5.0	794	2	S61698	hypothetical protein	7.87e+02
705	5	5.0	684	2	I39595	phospholipase C - Aer	7.87e+02	778	5	5.0	794	2	S06687	hypothetical protein	7.87e+02
706	5	5.0	684	2	H71237	hypothetical protein	7.87e+02	779	5	5.0	796	2	S66769	probable membrane pro	7.87e+02
707	5	5.0	691	2	S41008	hypothetical protein	7.87e+02	780	5	5.0	797	2	S44546	probable membrane pro	7.87e+02
708	5	5.0	692	2	T12587	NADH dehydrogenase su	7.87e+02	781	5	5.0	797	2	B28551	levansucrase (EC 2.4.	7.87e+02
709	5	5.0	694	2	T07638	probable SWH1 protein	7.87e+02	782	5	5.0	800	2	S33079	PEF111 protein - yeas	7.87e+02
710	5	5.0	698	2	T12563	NADH dehydrogenase su	7.87e+02	783	5	5.0	802	2	S25957	gene cox1 intron 2 pr	7.87e+02
711	5	5.0	698	2	T12556	NADH dehydrogenase su	7.87e+02	784	5	5.0	803	2	S45916	hypothetical protein	7.87e+02
712	5	5.0	698	2	T12562	NADH dehydrogenase su	7.87e+02	785	5	5.0	807	2	A64575	VirB4 homolog - Helic	7.87e+02
713	5	5.0	698	2	T12560	NADH dehydrogenase su	7.87e+02	786	5	5.0	808	2	T14513	hypothetical protein	7.87e+02
714	5	5.0	698	2	T12569	NADH dehydrogenase su	7.87e+02	787	5	5.0	821	2	S39983	eps8 protein - mouse	7.87e+02
715	5	5.0	698	2	T12586	NADH dehydrogenase su	7.87e+02	788	5	5.0	822	2	I38728	epidermal growth fact	7.87e+02
716	5	5.0	698	2	T12588	NADH dehydrogenase su	7.87e+02	789	5	5.0	830	2	B44439	protein kinase (EC 2.	7.87e+02
717	5	5.0	698	2	T12590	NADH dehydrogenase su	7.87e+02	790	5	5.0	831	2	T05265	coat protein gamma-CO	7.87e+02
718	5	5.0	698	2	T12564	NADH dehydrogenase su	7.87e+02	791	5	5.0	834	2	S66498	M-sema F protein prec	7.87e+02
719	5	5.0	698	2	T12565	NADH dehydrogenase su	7.87e+02	792	5	5.0	835	1	S54216	invasin - Yersinia en	7.87e+02
720	5	5.0	700	2	T13763	NADH dehydrogenase su	7.87e+02	793	5	5.0	835	2	I55603	reduced hepatic gluc	7.87e+02
721	5	5.0	700	2	F64078	translation elongatio	7.87e+02	794	5	5.0	839	2	B64689	site-specific DNA-met	7.87e+02
722	5	5.0	702	2	S65544	CAMP-dependent phosph	7.87e+02	795	5	5.0	839	2	S73548	MG422 homolog C12-orf	7.87e+02
723	5	5.0	704	1	EFECC	translation elongatio	7.87e+02	796	5	5.0	843	2	S38364	membrane alanyl amino	7.87e+02
724	5	5.0	704	2	T02902	hypothetical protein	7.87e+02	797	5	5.0	843	1	JDVVLJ3	DNA-directed DNA poly	7.87e+02
725	5	5.0	704	2	JC1424	translation elongatio	7.87e+02	798	5	5.0	845	2	H71317	probable methyl-accept	7.87e+02
726	5	5.0	707	2	S68858	finger protein - mous	7.87e+02	799	5	5.0	845	1	S34027	replication licensing	7.87e+02
727	5	5.0	710	2	H44490	retrovirus-related re	7.87e+02	800	5	5.0	848	4	A44282	retrovirus-related po	7.87e+02
728	5	5.0	711	2	S68443	double-stranded RNA-s	7.87e+02	801	5	5.0	848	2	T16430	hypothetical protein	7.87e+02
729	5	5.0	711	2	A72375	hypothetical protein	7.87e+02	802	5	5.0	852	1	VCLJGG	env polypeptide precu	7.87e+02
730	5	5.0	713	2	JC6084	lambda-interacting pr	7.87e+02	803	5	5.0	853	2	T08162	amylopullulanase (EC	7.87e+02
731	5	5.0	713	2	S64437	probable membrane pro	7.87e+02	804	5	5.0	855	2	S46050	probable purine nucle	7.87e+02
732	5	5.0	716	2	D71631	hypothetical protein	7.87e+02	805	5	5.0	856	2	A44439	protein kinase (EC 2.	7.87e+02
733	5	5.0	716	2	H72635	hypothetical protein	7.87e+02	806	5	5.0	857	2	A42861	protein kinase TTK (E	7.87e+02
734	5	5.0	716	3	T10932	guanosine pentaphosph	7.87e+02	807	5	5.0	860	2	T15778	hypothetical protein	7.87e+02
735	5	5.0	719	2	JC2181	differentiation-stimu	7.87e+02	808	5	5.0	865	2	D72206	valine--tRNA ligase (7.87e+02
736	5	5.0	719	2	S51739	transcription repress	7.87e+02	809	5	5.0	869	2	C56617	cfac protein precurs	7.87e+02
737	5	5.0	724	2	S57604	probable membrane pro	7.87e+02	810	5	5.0	877	2	S49783	probable membrane pro	7.87e+02
738	5	5.0	725	2	T10464	fatty acid oxidation	7.87e+02	811	5	5.0	878	2	S74207	lipoygenase (EC 1.13	7.87e+02
739	5	5.0	727	2	A58457	segment S4 protein -	7.87e+02	812	5	5.0	880	1	VCLJS2	env polypeptide precu	7.87e+02
740	5	5.0	727	1	A45342	nonstructural protein	7.87e+02	813	5	5.0	881	2	S56032	probable membrane pro	7.87e+02
741	5	5.0	728	2	F72693	hypothetical protein	7.87e+02	814	5	5.0	882	2	A42855	N-heparan sulfate sul	7.87e+02
742	5	5.0	730	1	T7HUSK	transforming protein	7.87e+02	815	5	5.0	884	1	IJMSCE	E-cadherin precursor,	7.87e+02
743	5	5.0	728	2	S28294	hypothetical protein	7.87e+02	816	5	5.0	885	2	S67660	hypothetical protein	7.87e+02
744	5	5.0	731	2	T01302	hypothetical protein	7.87e+02	817	5	5.0	892	2	JQ0424	probable transposase	7.87e+02
745	5	5.0	733	2	S31288	MAK10 protein - yeast	7.87e+02	818	5	5.0	900	2	S70630	xeroderma pigmentosum	7.87e+02
746	5	5.0	742	1	S38691	kinesin-related prote	7.87e+02	819	5	5.0	902	2	T16821	hypothetical protein	7.87e+02
747	5	5.0	742	2	S55098	probable membrane pro	7.87e+02	820	5	5.0	906	2	S25852	glutamate receptor GL	7.87e+02
748	5	5.0	744	2	A43353	ascites sialoglycopro	7.87e+02	821	5	5.0	906	2	S38723	glutamate receptor GL	7.87e+02
749	5	5.0	744	2	A45622	surface antigen gp85	7.87e+02	822	5	5.0	911	2	A56465	TFIIIC2 subunit - hum	7.87e+02
750	5	5.0	745	2	G72453	hypothetical protein	7.87e+02	823	5	5.0	911	2	S51441	hypothetical protein	7.87e+02
751	5	5.0	747	1	VPXRA5	outer layer protein v	7.87e+02	824	5	5.0	914	2	T00757	hypothetical protein	7.87e+02
752	5	5.0	753	2	C71086	hypothetical protein	7.87e+02	825	5	5.0	914	2	B48086	translation initiatio	7.87e+02
753	5	5.0	756	2	A43582	surface antigen mspl	7.87e+02	826	5	5.0	916	2	A38418	jockey protein 2 - fr	7.87e+02

827	5	5.0	916	2	JT0396	reverse transcriptase	7.87e+02	900	2	A40670	nuclear envelope prot	7.87e+02
828	5	5.0	919	1	RNVZCA	DNA-directed RNA poly	7.87e+02	901	2	S71424	nitric-oxide synthase	7.87e+02
829	5	5.0	920	2	A36887	endo-alpha-sialidase	7.87e+02	902	2	JW0105	synaptotagmin 2 alpha	7.87e+02
830	5	5.0	921	1	PNXNCP	H+-transporting ATPase	7.87e+02	903	2	T14265	golgin-245 - mouse	7.87e+02
831	5	5.0	920	2	G71705	alkaline phosphatase	7.87e+02	904	2	T14157	serine/threonine prot	7.87e+02
832	5	5.0	932	2	S09151	suvar(3'7) protein - f	7.87e+02	905	2	T01956	hypothetical protein	7.87e+02
833	5	5.0	939	2	T05209	hypothetical protein	7.87e+02	906	2	T14270	hypothetical protein	7.87e+02
834	5	5.0	946	2	S64745	Sfil protein - yeast	7.87e+02	907	2	T14349	Ras-GTPase activating	7.87e+02
835	5	5.0	952	1	A32609	alpha-glucosidase (EC	7.87e+02	908	2	S62544	hypothetical protein	7.87e+02
836	5	5.0	954	2	S57108	hypothetical protein	7.87e+02	909	2	T00443	hypothetical protein	7.87e+02
837	5	5.0	956	2	S30834	hypothetical protein	7.87e+02	910	2	T09204	probable tail-host sp	7.87e+02
838	5	5.0	961	2	T01167	hypothetical protein	7.87e+02	911	2	A42150	p-glycoprotein pppl -	7.87e+02
839	5	5.0	961	1	PL1BVA	RNA Ia protein - brom	7.87e+02	912	2	T16507	hypothetical protein	7.87e+02
840	5	5.0	962	2	T16167	SERA antigen/papain-1	7.87e+02	913	2	B71413	hypothetical protein	7.87e+02
841	5	5.0	964	2	S45944	hypothetical protein	7.87e+02	914	2	B53739	vacuolating cytotoxin	7.87e+02
842	5	5.0	973	2	S54530	coatomer complex beta	7.87e+02	915	2	E71884	vacuolating cytotoxin	7.87e+02
843	5	5.0	983	2	H72510	probable ribonucleoti	7.87e+02	916	2	T09273	probable tail-host sp	7.87e+02
844	5	5.0	985	2	E69850	formate dehydrogenase	7.87e+02	917	2	T14259	ras GTPase-activating	7.87e+02
845	5	5.0	987	2	I4B373	G-utrophin - mouse	7.87e+02	918	2	A54831	nuclear pore complex	7.87e+02
846	5	5.0	991	1	PL1BVCV	RNA I protein - cucum	7.87e+02	919	2	Q0BE1	membrane antigen p140	7.87e+02
847	5	5.0	992	2	S04781	exonuclease ABC chal	7.87e+02	920	2	T14316	rig-1 protein - mouse	7.87e+02
848	5	5.0	997	2	F71365	probable DNA polymera	7.87e+02	921	2	JC4953	vascular endothelial	7.87e+02
849	5	5.0	999	2	F72453	probable cytochrome C	7.87e+02	922	2	T10803	probable RNA-directed	7.87e+02
850	5	5.0	1002	2	A56678	yeamanuclein-alpha - f	7.87e+02	923	2	S51389	ROM2 protein - yeast	7.87e+02
851	5	5.0	1004	2	A71617	SERA antigen/papain-1	7.87e+02	924	2	T12064	DNA binding protein G	7.87e+02
852	5	5.0	1008	2	T12532	hypothetical protein	7.87e+02	925	2	A29959	DNA-directed RNA poly	7.87e+02
853	5	5.0	1011	2	T13669	neuromusculin - fruit	7.87e+02	926	2	A75207	amylopullulanase PAB0	7.87e+02
854	5	5.0	1011	2	T13055	dynamitin associated pr	7.87e+02	927	2	JC5778	apoptosis signal-regu	7.87e+02
855	5	5.0	1014	2	S75724	hypothetical protein	7.87e+02	928	2	T13052	guanine nucleotide ex	7.87e+02
856	5	5.0	1017	2	D75028	h+-transporting ATP s	7.87e+02	929	2	A43336	microtubule-vesicle 1	7.87e+02
857	5	5.0	1034	2	S63536	aminomethyltransferas	7.87e+02	930	2	A29637	position-specific ant	7.87e+02
858	5	5.0	1035	2	D71831	cation efflux system	7.87e+02	931	2	S25997	gene atpA intron 1 pr	7.87e+02
859	5	5.0	1035	1	A43090	enteropeptidase (EC 3	7.87e+02	932	2	B42239	adenylate cyclase (EC	7.87e+02
860	5	5.0	1036	2	A42895	H+/K+-exchanging ATPa	7.87e+02	933	2	A57013	early endosome antige	7.87e+02
861	5	5.0	1037	2	S40216	P protein - Flaveria	7.87e+02	934	2	T01610	RNA-directed DNA poly	7.87e+02
862	5	5.0	1037	2	S63535	aminomethyltransferas	7.87e+02	935	2	A57570	Bloom's syndrome rela	7.87e+02
863	5	5.0	1038	2	T15098	hypothetical protein	7.87e+02	936	2	S22695	restin - human	7.87e+02
864	5	5.0	1044	2	T02615	probable glycine dehy	7.87e+02	937	2	BVBYL1	guanine nucleotide-re	7.87e+02
865	5	5.0	1045	2	JC5795	CDEP protein - human	7.87e+02	938	2	B71610	WD40 WEB-1 homolog PF	7.87e+02
866	5	5.0	1048	2	A70592	hypothetical protein	7.87e+02	939	2	A45665	adult-specific brush	7.87e+02
867	5	5.0	1050	2	JW0092	serine-threonine kina	7.87e+02	940	2	A36426	SPA2 protein - yeast	7.87e+02
868	5	5.0	1052	2	T14343	zinc finger RNA bindi	7.87e+02	941	2	A47341	amylopullulanase prec	7.87e+02
869	5	5.0	1053	2	T09645	DNA-directed RNA poly	7.87e+02	942	2	T02850	probable membrane pro	7.87e+02
870	5	5.0	1053	2	T09641	DNA-directed RNA poly	7.87e+02	943	2	T05634	hypothetical protein	7.87e+02
871	5	5.0	1054	2	B38919	hypothetical protein	7.87e+02	944	2	A40228	neurexin I-alpha prec	7.87e+02
872	5	5.0	1057	2	T10908	DNA-directed RNA poly	7.87e+02	945	2	T14961	hypothetical protein	7.87e+02
873	5	5.0	1058	2	S50295	hypothetical protein	7.87e+02	946	2	A44406	DNA topoisomerase (AT	7.87e+02
874	5	5.0	1058	2	S63460	apolipoprotein B - At	7.87e+02	947	2	JS0703	DNA topoisomerase (AT	7.87e+02
875	5	5.0	1062	2	T14151	inv protein - mouse	7.87e+02	948	2	I45944	neurexin I-alpha - bo	7.87e+02
876	5	5.0	1062	2	S09834	hypothetical protein	7.87e+02	949	2	A40493	DNA topoisomerase (AT	7.87e+02
877	5	5.0	1077	2	S45395	homeotic protein PRHP	7.87e+02	950	2	A56734	ribosome receptor, 18	7.87e+02
878	5	5.0	1088	2	T14917	isolectine-tRNA liga	7.87e+02	951	2	JC4172	DNA (cytosine-5-)-met	7.87e+02
879	5	5.0	1091	2	E71322	differentiation-stimu	7.87e+02	952	2	G70630	probable ctPH protein	7.87e+02
880	5	5.0	1092	2	JX0312	dynamitin associated pr	7.87e+02	953	2	VGWJBV	peplomeric glycoprotein	7.87e+02
881	5	5.0	1094	2	T13053	surface-array protein	7.87e+02	954	2	S48933	probable transport pr	7.87e+02
882	5	5.0	1109	2	A56143	G2-G1 polyprotein - I	7.87e+02	955	2	A41216	guanine nucleotide ex	7.87e+02
883	5	5.0	1110	1	B42544	NF-180 - sea lamprey	7.87e+02	956	2	T08880	NMDA receptor-binding	7.87e+02
884	5	5.0	1110	2	I51116	RNA-splicing regulato	7.87e+02	957	2	RRGCMV	genome polyprotein -	7.87e+02
885	5	5.0	1117	2	A38227	pol polyprotein - fel	7.87e+02	958	2	S53457	dominant autoantigen	7.87e+02
886	5	5.0	1124	2	S23820	pol polyprotein - fel	7.87e+02	959	2	JC1340	outer membrane protei	7.87e+02
887	5	5.0	1124	1	GNLJFP	pol polyprotein - fel	7.87e+02	960	2	T13748	sex comb protein - fr	7.87e+02
888	5	5.0	1143	2	S46122	SNF2 protein homolog	7.87e+02	961	2	S77908	hypothetical protein	7.87e+02
889	5	5.0	1144	2	A54810	TMV resistance protei	7.87e+02	962	2	C40228	neurexin II-alpha pre	7.87e+02
890	5	5.0	1144	2	A75132	hypothetical protein	7.87e+02	963	2	T14867	interaptin - slime mo	7.87e+02
891	5	5.0	1144	2	S57348	nuclear factor Rpl140	7.87e+02	964	2	S57329	ubiquitin-protein lig	7.87e+02
892	5	5.0	1158	2	I59311	NRD convertase - rat	7.87e+02	965	2	T14106	tuberosin sclerosits 2	7.87e+02
893	5	5.0	1161	2	I59311	synGAP-bi protein - r	7.87e+02	966	2	B70984	probable GTPase-activ	7.87e+02
894	5	5.0	1166	2	T13958	chromosome segregatio	7.87e+02	967	2	S48938	hypothetical protein	7.87e+02
895	5	5.0	1177	2	I49017	calcium-activated pot	7.87e+02	968	2	T03224	probable polyketide s	7.87e+02
896	5	5.0	1184	2	I49017	type I site-specific	7.87e+02	969	2	T13937	hypothetical protein	7.87e+02
897	5	5.0	1188	1	NDECKR	Pm5 protein - human	7.87e+02	970	2	T16871	RNA-directed RNA poly	7.87e+02
898	5	5.0	1190	2	S21977	calcium-activated pot	7.87e+02	971	2	T08212		
899	5	5.0	1196	2	A48206			972	2			

RESULT	2	#type complete
ENTRY	A30350	

RESULT	2	#type complete
ENTRY	A30350	

RESULT	2	#type complete
ENTRY	A30350	

Best Local Similarity 100.0%; Pred. No. 2.34e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 65 NLDHIA 71
|||||
Qy 86 NLDHIA 92

RESULT 4

ENTRY A70666 #type complete
TITLE probable modA protein - Mycobacterium tuberculosis (strain H37Rv)

ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 26-Aug-1999

ACCESSIONS A70666

REFERENCE A70500

#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

#journal Nature (1998) 393:537-544

#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.

#cross-references MUID:98295987

#accession A70666

##status preliminary; nucleic acid sequence not shown;

##molecule_type DNA

##residues 1-261 ##label COL

##cross-references GB:283859; GB:AL123456; NID:g3261678; PID:e290724; PID:g1781190

##experimental_source strain H37Rv

GENETICS

#gene modA

CLASSIFICATION #superfamily molybdate-binding periplasmic protein

SUMMARY #length 261 #molecule-weight 26576 #checksum 3096

Query Match 6.9%; Score 7; DB 2; Length 261;
Best Local Similarity 100.0%; Pred. No. 2.34e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

74 AGSSELA 80
|||||
Qy 64 AGSSELA 70

RESULT 5

ENTRY B53419 #type complete

TITLE biphenyl-2,3-diol 1,2-dioxygenase (EC 1.13.11.39) I -

ALTERNATE_NAMES Rhodococcus globerulus

ORGANISM 2,3-dihydroxybiphenyl dioxygenase I

DATE #formal_name Rhodococcus globerulus

25-May-1994 #sequence_revision 06-Jan-1995 #text_change 11-Jun-1999

ACCESSIONS B53419

REFERENCE A53419

#authors Asturias, J.A.; Eltis, L.D.; Prucha, M.; Timmis, K.N.

#journal J. Biol. Chem. (1994) 269:7807-7815

#title Analysis of three 2,3-dihydroxybiphenyl 1,2-dioxygenases

found in Rhodococcus globerulus P6. Identification of a new

family of extradiol dioxygenases.

#cross-references MUID:94171820

#accession B53419

##molecule_type DNA

##residues 1-291 ##label AST

##cross-references GB:X75633; NID:g473115; PIDN:CAA53297.1; PID:g473117

##experimental_source strain P6

COMMENT This enzyme catalyzes the third step in the major degradative pathway for biphenyl and polychlorinated biphenyls (PCBs): cleavage of a 2,3-dihydroxybiphenyl derivative at the 1 and 2 positions to give a derivative of 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate.

COMMENT This is one of the three biphenyl-2,3-diol 1,2-dioxygenases found in *R. globerulus* P6; it is the only one that exhibits homology with the corresponding proteins of analogous degradative pathways in gram-negative bacteria. The other two, biphenyl-2,3-diol 1,2-dioxygenases II and III, although homologous, belong to another class of biphenyl-2,3-diol 1,2-dioxygenases.

GENETICS

#gene bphC1

CLASSIFICATION #superfamily biphenyl-2,3-diol 1,2-dioxygenase

KEYWORDS aromatic hydrocarbon catabolism; iron; oxidoreductase; PCB biodegradation

SUMMARY #length 291 #molecule-weight 32081 #checksum 1071

Query Match 6.9%; Score 7; DB 1; Length 291;

Best Local Similarity 100.0%; Pred. No. 2.34e-01;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 91 SSELAAE 97

Qy 66 SSELAAE 72

|||||

RESULT 6

ENTRY S61646 #type complete

TITLE dolichyl-diphosphooligosaccharide--protein glycotransferase (EC 2.4.1.119) gamma chain - yeast (*Saccharomyces cerevisiae*)

ALTERNATE_NAMES protein O3124; protein YOR085w; protein YOR31124w

ORGANISM #formal_name *Saccharomyces cerevisiae*

DATE 09-Mar-1996 #sequence_revision 12-Apr-1996 #text_change 07-May-1999

ACCESSIONS S61646; S64645; S64653; S66970

REFERENCE S61643

#authors Benes, V.; Andrade, M.A.; Rechmann, S.; Teodoru, C.; Banrevi, A.; Sander, C.; Valencia, A.; Ansorge, W.; Voss, H.

#submission submitted to the EMBL Data Library, December 1995

#description Nucleotide sequence and analysis of a 130 kb fragment of yeast chromosome XV.

#accession S61646

##molecule_type DNA

##residues 1-350 ##label BEN

##cross-references EMBL:X94335; NID:gl262139; PID:e217723; PID:gl164932

REFERENCE S64645

#authors Karacoglou, D.; Kelleher, D.J.; Gilmore, R.

#journal J. Cell Biol. (1995) 130:567-577

#title Functional characterization of Ost3p. Loss of the 34-kD subunit of the *Saccharomyces cerevisiae* oligosaccharyltransferase results in biased underglycosylation of acceptor substrates.

#cross-references MUID:95348180

#accession S64645

##molecule_type DNA

##residues 1-350 ##label KAR

##cross-references EMBL:U25052; NID:g951121; PID:g951122

REFERENCE S64653

#accession S64653

##molecule_type protein

##residues 23-37;99-103;123-153 ##label KAZ

REFERENCE S66965

#authors Voss, H.; Benes, V.; Rechmann, S.; Teodoru, C.; Schwager, C.;

Paces, V.; Ansorge, W.

#submission submitted to the Protein Sequence Database, July 1996

#accession S66970

##molecule_type DNA

##residues 1-350 ##label VOS

##cross-references EMBL:274993; NID:gl420249; PID:e252347; PID:gl420250; MIPS:YOR085w

##experimental_source strain S288C

```

GENETICS
#gene SGD:OST3
##cross-references SGD:S0005611; MIPS:YOR085W
#map_position 15R
KEYWORDS glycosyltransferase; hexosyltransferase; transmembrane protein
FEATURE
1-17 #domain transmembrane #status predicted #label TM1\
187-203 #domain transmembrane #status predicted #label TM2\
218-234 #domain transmembrane #status predicted #label TM3\
272-288 #domain transmembrane #status predicted #label TM4\
311-327 #domain transmembrane #status predicted #label TM5\
SUMMARY #length 350 #molecular-weight 39483 #checksum 9523

Query Match 6.9%; Score 7; DB 2; Length 350;
Best Local Similarity 100.0%; Pred. No. 2.34e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 241 RNTQLAG 247
( 59 RNTQLAG 65

RESULT 7
ENTRY E71912 #type complete
TITLE polynucleotide adenylyltransferase - Helicobacter pylori
ORGANISM #formal_name Helicobacter pylori
#variety strain J99
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
12-Feb-1999
E71912
A71800
Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
G.F.; Trust, T.J.
#journal Nature (1999) 397:176-180
#title Genomic sequence comparison of two unrelated isolates of the
human gastric pathogen Helicobacter pylori.
#cross-references MUID:99120557
#status preliminary
#residues 1-402 #label ARN
#experimental_source strain J99
#gene pcnB
SUMMARY #length 402 #molecular-weight 46425 #checksum 1455

Query Match 6.9%; Score 7; DB 2; Length 402;
Best Local Similarity 100.0%; Pred. No. 2.34e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 364 LNSVNIEN 370
( 36 LNSVNIEN 42

RESULT 8
ENTRY B71858 #type complete
TITLE adenylosuccinate lyase - Helicobacter pylori (strain J99)
ORGANISM #formal_name Helicobacter pylori
#variety strain J99
DATE 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change
22-Jun-1999
B71858
A71800
Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.;
Doig, P.C.; Smith, D.R.; Noonan, B.; Guild, B.C.; deJonge,
B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
G.F.; Trust, T.J.
#journal Nature (1999) 397:176-180
#title Genomic sequence comparison of two unrelated isolates of the
human gastric pathogen Helicobacter pylori.
#cross-references MUID:99120557
#status preliminary
#residues 1-402 #label ARN
#experimental_source strain J99
#gene pcnB
SUMMARY #length 402 #molecular-weight 46425 #checksum 1455

GENETICS
#gene xynF
FUNCTION #description catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in
xylans
#pathway xylan degradation
CLASSIFICATION #superfamily Pseudomonas endo-1,4-beta-xylanase F;
Streptomyces endo-1,4-beta-xylanase A homology
glycosidase; hydrolase; polysaccharide degradation
KEYWORDS #domain signal sequence #status predicted #label SIG\
#product endo-1,4-beta-xylanase F #status predicted
#label MAT\
#domain Streptomyces endo-1,4-beta-xylanase A homology
#active_site Glu #status predicted
#label SXV\
SUMMARY #length 606 #molecular-weight 64841 #checksum 6457

Query Match 6.9%; Score 7; DB 1; Length 606;
Best Local Similarity 100.0%; Pred. No. 2.34e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

B.L.; Carmel, G.; Tummino, P.J.; Caruso, A.;
Uria-Nickelsen, M.; Mills, D.M.; Ives, C.; Gibson, R.;
Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis,
G.F.; Trust, T.J.
#journal Nature (1999) 397:176-180
#title Genomic sequence comparison of two unrelated isolates of the
human gastric pathogen Helicobacter pylori.
#cross-references MUID:99120557
#accession B71858
#status preliminary
#molecule_type DNA
#residues 1-440 #label ARN
#cross-references GB:AE001531; GB:AE001439; NID:g4155617;
PIDN:AA06609.1; PID:g4155619
#experimental_source strain J99

```

```

GENETICS
#gene purB
CLASSIFICATION #superfamily fumarate hydratase
SUMMARY #length 440 #molecular-weight 49867 #checksum 8630

```

```

Query Match 6.9%; Score 7; DB 2; Length 440;
Best Local Similarity 100.0%; Pred. No. 2.34e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 327 LNSVNIEN 333
( 36 LNSVNIEN 42

```

```

RESULT 9
ENTRY S59634 #type complete
TITLE endo-1,4-beta-xylanase (EC 3.2.1.8) F precursor - Pseudomonas
fluorescens
ORGANISM #formal_name Pseudomonas fluorescens
DATE 15-Feb-1996 #sequence_revision 11-Apr-1997 #text_change
18-Jun-1999
S59634; S52749
S59631
#authors Millward-sadler, S.J.; Davidson, K.; Hazlewood, G.P.; Black,
G.W.; Gilbert, H.J.; Clarke, J.H.
#journal Biochem. J. (1995) 312:39-48
#title Novel cellulose-binding domains, NodB homologues and
conserved modular architecture in xylanases from the
aerobic soil bacteria Pseudomonas fluorescens subsp.
cellulosa and Cellvibrio mixtus.
#cross-references MUID:96077124
#accession S59634
#molecule_type DNA
#residues 1-606 #label MIL
#cross-references EMBL:248928; NID:g758226; PIDN:CAA88764.1;
PID:g758227

```

```

GENETICS
#gene xynF
FUNCTION #description catalyzes the hydrolysis of 1,4-beta-xylosidic linkages in
xylans
#pathway xylan degradation
CLASSIFICATION #superfamily Pseudomonas endo-1,4-beta-xylanase F;
Streptomyces endo-1,4-beta-xylanase A homology
glycosidase; hydrolase; polysaccharide degradation
KEYWORDS #domain signal sequence #status predicted #label SIG\
#product endo-1,4-beta-xylanase F #status predicted
#label MAT\
#domain Streptomyces endo-1,4-beta-xylanase A homology
#active_site Glu #status predicted
#label SXV\
SUMMARY #length 606 #molecular-weight 64841 #checksum 6457

```

```

Query Match 6.9%; Score 7; DB 1; Length 606;
Best Local Similarity 100.0%; Pred. No. 2.34e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Db 184 LAGSSEL 190
QY 63 LAGSSEL 69
|||||
QY 64 AGSSELA 70

RESULT 10
ENTRY H64888 #type complete
TITLE membrane protein ydbH - Escherichia coli
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 13-Nov-1998
ACCESSIONS H64888
REFERENCE A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession H64888
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-879 ##label BLAT
##cross-references GB:AE000235; GB:U00096; NID:gl787643; PID:gl787646; UWGP:bl381
#experimental_source strain K-12, substrain MG1655
GENETICS ydbH
#gene
FEATURE 8-24
SUMMARY #domain transmembrane #status predicted #label TM01
#length 879 #molecular-weight 96834 #checksum 5911
Query Match 6.9%; Score 7; DB 2; Length 879;
Best Local Similarity 100.0%; Pred. No. 2.34e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 801 STKINLD 807
QY 82 STKINLD 88
|||||

RESULT 11
ENTRY T14260 #type complete
TITLE period protein Per3 - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
ACCESSIONS T14260
REFERENCE T17943
#authors Zylka, M.J.; Shearman, L.P.; Weaver, D.R.; Reppert, S.M.
#journal Neuron (1998) 20:1103-1110
#title Three period homologs in mammals: differential light responses in the suprachiasmatic circadian clock and oscillating transcripts outside of brain.
#cross-references MUID:98318231
#accession T14260
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type mRNA
#residues 1-1113 ##label 2YL
##cross-references EMBL:AF050182; NID:g3136149; PID:g3136150; PIDN:AAC40147.1
GENETICS Per3
KEYWORDS circadian rhythm
SUMMARY #length 1113 #molecular-weight 120938 #checksum 5672
Query Match 6.9%; Score 7; DB 2; Length 1113;
Best Local Similarity 100.0%; Pred. No. 2.34e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 863 AGSSELA 869
QY 64 AGSSELA 70
|||||

RESULT 12
ENTRY T13955 #type complete
TITLE period protein Per3 - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
ACCESSIONS T13955
REFERENCE T17831
#authors Takumi, T.; Taguchi, K.; Miyake, S.; Sakakida, Y.; Takashima, N.; Matsubara, C.; Maebayashi, Y.; Okumura, K.; Takekida, S.; Yamamoto, S.; Yagita, K.; Yan, L.; Young, M.W.; Okamura, H.
#journal EMBO J. (1998) 17:4753-4759
#title A light-independent oscillatory gene mPer3 in mouse SCN and OVLT.
#cross-references MUID:98372739
#accession T13955
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type mRNA
#residues 1-1115 ##label TAK
##cross-references EMBL:AB013605; NID:dl229778; PID:dl034431; PIDN:BAA33465.1
GENETICS Per3
#gene
SUMMARY #length 1115 #molecular-weight 121040 #checksum 1988
Query Match 6.9%; Score 7; DB 2; Length 1115;
Best Local Similarity 100.0%; Pred. No. 2.34e-01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 863 AGSSELA 869
QY 64 AGSSELA 70
|||||

RESULT 13
ENTRY SNPSO #type complete
TITLE ice nucleation protein - Pseudomonas syringae
ALTERNATE_NAMES Snomax (TM)
ORGANISM #formal_name Pseudomonas syringae
DATE 31-Mar-1988 #sequence_revision 21-Jan-1997 #text_change 16-Jul-1999
ACCESSIONS A24405
REFERENCE A24405
#authors Green, R.L.; Warren, G.J.
#journal Nature (1985) 317:645-648
#title Physical and functional repetition in a bacterial ice nucleation gene.
#accession A24405
#molecule_type DNA
#residues 1-1200 ##label GRE
##cross-references EMBL:X03035; NID:g45828; PIDN:CAA36837.1; PID:g45829
REFERENCE A51242
#authors Kajava, A.V.; Lindow, S.E.
#submission submitted to the Brookhaven Protein Data Bank, June 1993
#cross-references PDB:1INA
#contents annotation; theoretical model, residues 490-535
REFERENCE A58442
#authors Kajava, A.V.; Lindow, S.E.
#journal J. Mol. Biol. (1993) 232:709-717
#title A model of the three-dimensional structure of ice nucleation proteins.
#cross-references MUID:93360260
#contents annotation; theoretical model
COMMENT Found on the outer membrane of the bacteria, this protein stimulates ice formation and frost damage of fruit. It is manufactured from clones of Escherichia coli and used in snow making.
COMMENT Snomax is a trademark of Snomax Technologies, Rochester, NY.

```

CLASSIFICATION #superfamily ice nucleation protein
KEYWORDS tandem repeat
FEATURE
208-1151 #region 8-residue repeats (A-G-Y-G-S-T-L-T)
SUMMARY #length 1200 #molecular-weight 118587 #checksum 8220

Query Match 6.9%; Score 7; DB 1; Length 1200;
Best Local Similarity 100.0%; Pred. No. 2.34e+01;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 586 QGSVLTS 592
|||||
Qy 14 QGSVLTS 20

RESULT 14
ENTRY #type complete
TITLE hypothetical protein 252 - Autographa californica nuclear
ORGANISM polyhedrosis virus
#formal_name Autographa californica nuclear polyhedrosis
virus, ACMPV
20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
09-Sep-1997
ACCESSIONS S36691
REFERENCE S36690
#authors Kool, M.; Broer, R.; Zuidema, D.; Goldbach, R.W.; Vlak, J.M.
#submission submitted to the EMBL Data Library, April 1993
#description Nucleotide sequence of an Autographa californica nuclear
polyhedrosis virus 7.3 Kbp region (47 to 52.5 map units) of
ECORI-C.
#accession S36691
#status preliminary
#molecule_type DNA
#residues 1-84 #label KOO
#cross-references EMBL:X71415; NID:g525306; PID:g296322
SUMMARY #length 84 #molecular-weight 9513 #checksum 5616

Query Match 5.9%; Score 6; DB 2; Length 84;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 79 NTSSID 84
|||||
Qy 30 NTSSID 35

RESULT 15
ENTRY #type complete
TITLE ribosomal protein L36a.e.c8, cytosolic - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES protein YHR141c; ribosomal protein 44; ribosomal protein YL41
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 31-May-1979 #sequence_revision 12-May-1995 #text_change
22-Jun-1999
ACCESSIONS C43301; S48985; A02781
REFERENCE A43301
#authors Kawai, S.; Murao, S.; Mochizuki, M.; Shibuya, I.; Yano, K.;
Takagi, M.
#journal J. Bacteriol. (1992) 174:254-262
#title Drastic alteration of cycloheximide sensitivity by
substitution of one amino acid in the L41 ribosomal protein
of yeasts.
#cross-references MUID:92104971
#accession C43301
#molecule_type DNA
#residues 1-106 #label KAW
#cross-references EMBL:D10579; NID:g218483; PID:BAA01436.1;
PID:d1001910; PID:g218484
#note sequence extracted from NCBI backbone (NCBIN:75362,
NCBIP:75363)

REFERENCE S48966
#authors Fulton, L.
#submission submitted to the EMBL Data Library, June 1994

#description The sequence of S. cerevisiae cosmid 9315.
#accession S48985
#molecule_type DNA
#residues 1-106 #label FUL
#cross-references EMBL:U10398; NID:g551328; PIDN:AAB68420.1;
PID:g500688; MIPS:YHR141c
REFERENCE A02781
#authors T.; Wittmann-Liebold, B.
#journal FEBS Lett. (1978) 96:399-402
#title The primary structure of protein 44 from the large subunit of
yeast ribosomes.
#cross-references MUID:79086263
#accession A02781
#molecule_type protein
#residues 2-39,'RK',42-87,90-106 #label ITO
#note the residues at positions 40 and 54 were not positively
identified but are chemically related to
monomethyllysine

GENETICS SGD:MAK18
#gene #cross-references SGD:S0001183; MIPS:YHR141c
#map_position 8R
#introns 2/1
FUNCTION protein biosynthesis
#description protein biosynthesis
CLASSIFICATION #superfamily rat ribosomal protein L36a
KEYWORDS methylated amino acid; protein biosynthesis; ribosome
FEATURE
2-106 #product ribosomal protein L36a.e #status experimental
#label MATV
#modified_site lysine derivative (Lys) (probably
N6-methyllysine) #status experimental
SUMMARY #length 106 #molecular-weight 12212 #checksum 9122

Query Match 5.9%; Score 6; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 77 CKTRAQ 82
|||||
Qy 73 CKTRAQ 78

RESULT 16
ENTRY #type fragment
TITLE CD44 glycoprotein M1 isoform - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change
07-Feb-1997
ACCESSIONS A44355
REFERENCE A44355
#authors He, Q.; Lesley, J.; Hyman, R.; Ishihara, K.; Kincade, P.W.
#journal J. Cell Biol. (1992) 113:1711-1719
#title Molecular isoforms of murine CD44 and evidence that the
membrane proximal domain is not critical for hyaluronate
recognition.
#cross-references MUID:93107170
#accession A44355
#status preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues 1-107 #label HE1
#experimental_source squamous cell carcinoma KUN205
#note glycoprotein
#length 107 #checksum 6202

KEYWORDS
SUMMARY

Query Match 5.9%; Score 6; DB 2; Length 107;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 96 NVDGSL 101
|||||
Qy 42 NVDGSL 47

```

```

RESULT 17
ENTRY   #type complete
TITLE   hypothetical protein APE0660 - Aeropyrum pernix (strain K1)
ORGANISM #formal_name Aeropyrum pernix
DATE    20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change
ACCESSIONS
REFERENCE H72653
#authors Kawai, S.; Murao, S.; Mochizuki, H.; Yamazaki, S.;
          Haikawa, Y.; Jin-no, K.; Takahashi, M.; Sekine, M.; Baba,
          S.; Anka, A.; Kosugi, H.; Hosoyama, A.; Fukui, S.; Nagai,
          Y.; Nishijima, K.; Nakazawa, H.; Takamiya, M.; Masuda, S.;
          Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.;
          Kushiida, N.; Oguchi, A.; Aoki, K.; Kubota, K.; Nakamura,
          Y.; Nomura, N.; Sako, Y.; Kikuchi, H.
#journal DNA Res. (1999) 6:83-101
#title Complete genome sequence of an aerobic hyper-thermophilic
          Crenarchaeon, Aeropyrum pernix K1.
#cross-references MUID:99310339
#accession H72653
#status preliminary
#molecule_type DNA
#residues 1-109 #label KAW
#cross-references DDBJ:AP000060; NID:g5104188; PIDN:BAA79632.1;
          PID:d1043418; PID:g5104317
#experimental_source strain K1
GENETICS
#gene APE0660
#length 109 #molecular-weight 12078 #checksum 6545
Query Match 5.9%; Score 6; DB 2; Length 109;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 72 ENVDGS 77
Qy 41 ENVDGS 46

RESULT 18
ENTRY   #type complete
TITLE   ribosomal protein L36a.e.c14, cytosolic - yeast
          (Saccharomyces cerevisiae)
ALTERNATE_NAMES protein N1722; protein YNL162w; ribosomal protein YL41;
          ribosomal protein YL4a
ORGANISM #formal_name Saccharomyces cerevisiae
DATE    27-Apr-1996 #sequence_revision 03-May-1996 #text_change
ACCESSIONS S63114; S60965; B43301; S63812
REFERENCE S62967
#authors Nasr, F.; Becam, A.M.; Herbert, C.
#submission submitted to the Protein Sequence Database, April 1996
#accession S63114
#molecule_type DNA
#residues 1-116 #label NAS
#cross-references EMBL:Z71438; NID:g1302129; PIDN:CAA96049.1;
          PID:e239963; PID:g1302130; MIPS:YNL162w
#experimental_source strain S288C
REFERENCE S60958
#authors Nasr, F.; Becam, A.M.; Herbert, C.J.
#submission submitted to the EMBL Data Library, October 1995
#description The sequence of 36.8 kb from the left arm of chromosome XIV
          reveals 24 complete open reading frames; 18 correspond to
          new genes, one of which encodes a protein similar to the
          human myotonic dystrophy kinase.
#accession S60965
#molecule_type DNA
#residues 12-116 #label NAW
#cross-references EMBL:X92517; NID:g1050783; PIDN:CAA63277.1;
          PID:e206636; PID:g1326051
REFERENCE A43301

```

```

#authors Kawai, S.; Murao, S.; Mochizuki, M.; Shibuya, I.; Yano, K.;
          Takagi, M.
#journal J. Bacteriol. (1992) 174:254-262
#title Drastic alteration of cycloheximide sensitivity by
          substitution of one amino acid in the L41 ribosomal protein
          of yeasts.
#cross-references MUID:92104971
#accession B43301
#molecule_type DNA
#residues 'M', 12-116 #label KAW
#note sequence extracted from NCBI backbone (NCBIN:75360,
          NCBI:P:75361)
REFERENCE S63805
#authors Nasr, F.; Becam, A.M.; Herbert, C.J.
#journal Yeast (1996) 12:169-175
#title The sequence of 36.8 kb from the left arm of chromosome XIV
          reveals 24 complete open reading frames; 18 correspond to
          new genes, one of which encodes a protein similar to the
          human myotonic dystrophy kinase.
#cross-references MUID:96287653
#accession S63812
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 'M', 12-116 #label NAF
#cross-references EMBL:X92517; NID:g1050783; PIDN:CAA63277.1;
          PID:e206636; PID:g1326051
#note the nucleotide sequence was submitted to the EMBL Data
          Library, October 1995
GENETICS
#gene SGD:RPL41A; MAK18; SCL41A; YL41A
#cross-references SGD:S0005106; MIPS:YNL162w
#map_position 14L
#introns 2/1
CLASSIFICATION #superfamily_rat ribosomal protein L36a
KEYWORDS protein biosynthesis; ribosome
SUMMARY #length 116 #molecular-weight 13530 #checksum 1874
Query Match 5.9%; Score 6; DB 2; Length 116;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 87 CKTRAQ 92
Qy 73 CKTRAQ 78

RESULT 19
ENTRY   #type complete
TITLE   Ig V-D-J region (KR) - human
ORGANISM #formal_name Homo sapiens #common_name man
DATE    14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change
          23-Jul-1999
ACCESSIONS S69910
REFERENCE S69909
#authors Sahota, S.; Hamblin, T.; Oscier, D.G.; Stevenson, F.K.
#journal Leukemia (1994) 8:1285-1289
#title Assessment of the role of clonogenic B lymphocytes in the
          pathogenesis of multiple myeloma.
#cross-references MUID:94335315
#accession S69910
#status preliminary; nucleic acid sequence not shown
#molecule_type mRNA
#residues 1-122 #label SAH
#cross-references EMBL:Z33400; NID:g871350; PIDN:CAA83851.1;
          PID:g871351
#note the sequence of residues 108-122 and the corresponding
          nucleic acid sequence are not shown
CLASSIFICATION #superfamily immunoglobulin V region; immunoglobulin homology
SUMMARY #length 122 #molecular-weight 13511 #checksum 583
Query Match 5.9%; Score 6; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 100 SVLTST 105
QY 16 SVLTST 21

RESULT 20
ENTRY CD44 glycoprotein M2 isoform - mouse (fragment)
TITLE #formal_name Mus musculus #common_name house mouse
ORGANISM 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change
DATE 07-Feb-1997

ACCESSIONS B44355
REFERENCE He, Q.; Lesley, J.; Hyman, R.; Ishihara, K.; Kincade, P.W.
#authors J. Cell Biol. (1992) 119:1711-1719
#journal Molecular isoforms of murine CD44 and evidence that the
#title membrane proximal domain is not critical for hyaluronate
recognition.
#cross-references MUID:93107170
#accession B44355
#status preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues 1-142 #label HEL
#experimental_source squamous cell carcinoma KLN205
#note sequence extracted from NCBI backbone (NCBIP:121203)
KEYWORDS glycoprotein
SUMMARY #length 142 #checksum 7043

Query Match 5.9%; Score 6; DB 2; Length 142;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 131 NVDGSL 136
QY 42 NVDGSL 47

RESULT 21
ENTRY T04811 #type complete
TITLE STIG1 protein homolog F10M23.220 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
11-Jun-1999
T04811
ACCESSIONS Z15385
REFERENCE Bevan, M.; Lecharny, A.; Chefdor, F.; Krivitzky, M.; Kreis,
#authors M.; Hoheisel, J.; Mewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, February 1999
#accession T04811
#molecule_type DNA
#residues 1-152 #label BEV
#cross-references EMBL:AL035440
#experimental_source cultivar Columbia; BAC clone F10M23
GENETICS
#map_position 4
#note F10M23.220
SUMMARY #length 152 #molecular-weight 16540 #checksum 7330

Query Match 5.9%; Score 6; DB 2; Length 152;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 111 KFSQTC 116
QY 3 KFSQTC 8

RESULT 22
ENTRY I38108 #type complete
TITLE platelet-derived growth factor-BB - human
ORGANISM #formal_name Homo sapiens #common_name man

DATE 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change
16-Jul-1999
ACCESSIONS I38108
REFERENCE Cook, A.L.; Kirwin, P.M.; Craig, S.; Bawden, L.J.; Green,
#authors D.R.; Price, M.J.; Richardson, S.J.; Fallon, A.; Drummond,
A.H.; Edwards, R.M.; Clements, J.M.
Biochem. J. (1992) 281:57-65
#journal Purification and analysis of proteinase-resistant mutants of
#title recombinant platelet-derived growth factor-BB exhibiting
improved biological activity.
#cross-references MUID:92117992
#accession I38108
#status preliminary; translated from GB/EMBL/DDBJ
#molecule_type mRNA
#residues 1-161 #label RES
#cross-references EMBL:X63966; NID:g311378; PIDN:CAA45383.1; PID:g35377
CLASSIFICATION #superfamily platelet-derived growth factor
SUMMARY #length 161 #molecular-weight 18237 #checksum 8276

Query Match 5.9%; Score 6; DB 2; Length 161;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 15 AECKTR 20
QY 71 AECKTR 76

RESULT 23
ENTRY A46388 #type complete
TITLE Her-1 protein - Caenorhabditis elegans
ALTERNATE_NAMES masculinizing gene her-1
ORGANISM #formal_name Caenorhabditis elegans
DATE 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change
26-Aug-1999
ACCESSIONS A46388; S32245; S32246
REFERENCE Perry, M.D.; Li, W.; Trent, C.; Robertson, B.; Fire, A.;
#authors Hageman, J.M.; Wood, W.B.
Genes Dev. (1993) 7:216-228
#journal Molecular characterization of the her-1 gene suggests a
#title direct role in cell signaling during Caenorhabditis elegans
sex determination.
#cross-references MUID:93170661
#accession A46388
#status preliminary
#molecule_type DNA
#residues 1-175 #label PER
#cross-references EMBL:Z19595; NID:g297386; PID:g297387
#note Sequence extracted from NCBI backbone (NCBIN:125317,
NCBIN:125319, NCBIN:125321, NCBIP:125322)
#note the nucleotide sequence was submitted to the EMBL Data
Library, January 1993
GENETICS
#introns 48/2; 86/1; 138/3
CLASSIFICATION #superfamily Caenorhabditis elegans Her-1 protein
SUMMARY #length 175 #molecular-weight 20172 #checksum 5332

Query Match 5.9%; Score 6; DB 2; Length 175;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 81 KINLDD 86
QY 84 KINLDD 89

RESULT 24
ENTRY S75511 #type complete
TITLE hypothetical protein sll1239 - Synecocystis sp. (strain PCC
6803)
ORGANISM #formal_name Synecocystis sp.

```

Matches	6;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Db	69	SELAEE	74						
Qy	67	SELAEE	72						
RESULT	26								
ENTRY	WZVZAI	#type complete							
TITLE	21.7K HindIII-C protein - vaccinia virus (strain WR)								
ORGANISM	#formal_name vaccinia virus								
DATE	31-Mar-1990	#sequence_revision 31-Mar-1990	#text_change 16-Jul-1999						
ACCESSIONS	A31829								
REFERENCE	A94385								
#authors	Kotwal, G.J.; Moss, B.								
#journal	Virology (1988) 167:524-537								
#title	Analysis of a large cluster of nonessential genes deleted from a vaccinia virus terminal transposition mutant.								
#cross-references	MUID:89073756								
#accession	A31829								
#molecule_type	DNA								
#residues	1-184	#label KOT							
#cross-references	GB:M2812; NID:g335691; PIDN:AAA69592.1; PID:g893301								
CLASSIFICATION	#superfamily vaccinia virus 21.7K HindIII-C protein								
KEYWORDS	early protein								
SUMMARY	#length 184 #molecular-weight 21604 #checksum 2786								
Query Match	5.9%;	Score 6;	DB 1;	Length 184;					
Best Local Similarity	100.0%;	Pred. No. 1.85e+01;							
Matches	6;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Db	9	NTSSID	14						
Qy	30	NTSSID	35						
RESULT	27								
ENTRY	S58383	#type complete							
TITLE	hypothetical protein 2 - human								
ORGANISM	#formal_name Homo sapiens	#common_name man							
DATE	12-Feb-1998	#sequence_revision 20-Feb-1998	#text_change 16-Jul-1999						
ACCESSIONS	S58383								
REFERENCE	S58382								
#authors	Dicks, R.P.H.; Onnekink, C.; Jansen, H.J.; de Jong, A.;								
#journal	Bloemers, H.P.J.								
#title	Nucleic Acids Res. (1995) 23:2815-2822								
	A novel human c-sis mRNA species is transcribed from a promoter in c-sis intron 1 and contains the code for an alternative PDGF B-like protein.								
#cross-references	MUID:95388493								
#accession	S58383								
#status	preliminary								
#molecule_type	mRNA								
#residues	1-185	#label DIR							
#cross-references	EMBL:X83705; NID:g951023; PIDN:CAA58679.1;								
CLASSIFICATION	#superfamily platelet-derived growth factor								
SUMMARY	#length 185 #molecular-weight 20774 #checksum 2728								
Query Match	5.9%;	Score 6;	DB 2;	Length 185;					
Best Local Similarity	100.0%;	Pred. No. 1.85e+01;							
Matches	6;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Db	80	AECKTR	85						
Qy	71	AECKTR	76						
RESULT	28								
ENTRY	C69054	#type complete							
TITLE	fuculose-1-phosphate aldolase - Methanobacterium								


```

ORGANISM      thermoautotrophicum (strain Delta H)
#formal_name Methanobacterium thermoautotrophicum
#cross-references EMBL:567700; NID:g41754; PID:g41757
#accessions C69054
#reference A69000
#authors Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
Cook, R.; Gilbert, K.; Harrison, D.; Hoang, L.; Keagle, P.;
Lumm, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.;
Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.; Caruso,
A.; Bush, D.; Safer, H.; Patwell, D.; Prabhakar, S.;
McDougal, S.; Shimer, G.; Goyal, A.; Pietrovski, S.;
Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
J.; Reeve, J.N.
#journal J. Bacteriol. (1997) 179:7135-7155
#title Complete genome sequence of Methanobacterium
thermoautotrophicum Delta H: functional analysis and
comparative genomics.
#cross-references MUID:98037514
#accession C69054
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-191 #label MTH
#cross-references GB:AE000903; GB:AE000666; NID:g2622514; PID:g2622518
#experimental_source strain Delta H
GENETICS
#gene MTH1406
#start_codon GTG
#classification #superfamily Methanococcus jannaschii fucose-1-phosphate
aldolase
#summary #length 191 #molecular-weight 20702 #checksum 1063
Query Match 5.9%; Score 6; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 134 ELAAPC 139
|||||
Qy 68 ELAAPC 73
RESULT 29
ENTRY . #type complete
#title heat shock protein htpA - Escherichia coli
#classification #superfamily Escherichia coli heat shock protein htpA
#formal_name Escherichia coli
#organism Escherichia coli
#accessions A40623
#reference A40623; A56688; D64721; S28460
#authors Missiakas, D.; Georgopoulos, C.; Raina, S.
#journal J. Bacteriol. (1993) 175:2613-2624
#title The Escherichia coli heat shock gene htpA: mutational
analysis, cloning, sequencing, and transcriptional
regulation.
#cross-references MUID:93239687
#accession A40623
#status preliminary
#molecule_type DNA
#residues 1-196 #label MIS
#note sequence extracted from NCBI backbone (NCBIN:130422,
NCBIP:130423)
REFERENCE
#authors James, R.; Dean, D.O.; Debbage, J.
#journal DNA Seq. (1993) 3:327-332
#title Five open reading frames upstream of the dnaK gene of
Escherichia coli.
#cross-references MUID:94003405
#accession A56688
#status preliminary
#molecule_type DNA
thermoautotrophicum (strain Delta H)
#formal_name Methanobacterium thermoautotrophicum
#cross-references EMBL:567700; NID:g41754; PID:g41757
#accessions A64720
#reference A64720
#authors Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession D64721
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-196 #label BLAT
#cross-references GB:AE000112; GB:U00096; NID:gl786192; PID:gl786194;
#experimental_source strain K-12, substrain MG1655
#note it is uncertain whether Met-1 (CTG) or Met-36 is the
initiator
GENETICS
#gene htpA; htpY
#map_position 0 min
#start_codon CTG
#classification #superfamily Escherichia coli heat shock protein htpA
heat shock; stress-induced protein
#keywords #length 196 #molecular-weight 21225 #checksum 5329
#summary
Query Match 5.9%; Score 6; DB 2; Length 196;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 159 SIDLNS 164
|||||
Qy 33 SIDLNS 38
RESULT 30
ENTRY . #type complete
#title vif protein - bovine immunodeficiency virus (isolate 127)
#classification #superfamily vif protein
#formal_name bovine immunodeficiency virus
#organism bovine immunodeficiency virus
#accessions C34742
#reference C34742
#authors Garvey, K.J.; Oberste, M.S.; Elser, J.E.; Braun, M.J.; Gonda,
M.A.
#journal Virology (1990) 175:391-409
#title Nucleotide sequence and genome organization of biologically
active proviruses of the bovine immunodeficiency-like
virus.
#cross-references MUID:90223985
#accession C34742
#molecule_type genomic RNA
#residues 1-198 #label GAR
#cross-references GB:M32690; NID:g210706; PIDN:AAA91272.1; PID:g210709
GENETICS
#gene vif
#classification #superfamily BIV vif protein
#keywords AIDS; immunodeficiency
#summary #length 198 #molecular-weight 22828 #checksum 2644
Query Match 5.9%; Score 6; DB 1; Length 198;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 141 SVLTST 146
|||||
Qy 16 SVLTST 21
RESULT 31

```

```

ENTRY      A64124      #type complete
TITLE      GTP cyclohydrolase I (EC 3.5.4.16) - Haemophilus influenzae
            (strain Rd KW20)
ORGANISM   #formal_name Haemophilus influenzae
DATE       18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
            22-Jun-1999
ACCESSIONS A64124
REFERENCE   A64000
#authors   Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
            Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.;
            Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
            FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
            Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
            J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
            M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
            D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
            J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
            Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
#journal   Science (1995) 269:496-512
#title     Whole-genome random sequencing and assembly of Haemophilus
            influenzae Rd.
#cross-references MUID:95350630
#accession A64124
#status    nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues  1-218 #label TIGR
#cross-references GB:U32823; GB:I42023; NID:g1574281; PIDN:AAC23097.1;
            PID:g1574286; TIGR:HI1447
CLASSIFICATION #superfamily GTP cyclohydrolase I
KEYWORDS     hydrolase
SUMMARY      #length 218 #molecular-weight 24945 #checksum 3761

Query Match 5.9%; Score 6; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 105 LNSTCE 110
QY 18 LNSTCE 23

RESULT 32
ENTRY   D37141      #type complete
TITLE   resolvase - Escherichia coli plasmid RP4
ORGANISM #formal_name Escherichia coli
DATE    08-Mar-1991 #sequence_revision 08-Mar-1991 #text_change
            31-Oct-1997
ACCESSIONS D37141
REFERENCE   A37141
#authors   Gerlitz, M.; Hrabak, O.; Schwab, H.
            J. Bacteriol. (1990) 172:6194-6203
#journal   Partitioning of broad-host-range plasmid RP4 is a complex
            system involving site-specific recombination.
#cross-references MUID:91035226
#accession D37141
#status    preliminary
#molecule_type DNA
#residues  1-219 #label GER
#cross-references GB:M59825; GB:M37620; NID:g152539; PID:g152541
GENETICS
#genome   plasmid
KEYWORDS   DNA binding
SUMMARY    #length 219 #molecular-weight 24161 #checksum 5951

Query Match 5.9%; Score 6; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 116 SELAAE 121
QY 67 SELAAE 72

```

```

RESULT 33
ENTRY   D69778      #type complete
TITLE   hypothetical protein ydeJ - Bacillus subtilis
ORGANISM #formal_name Bacillus subtilis
DATE    05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
            24-Sep-1998
ACCESSIONS D69778
REFERENCE   A69580
#authors   Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
            Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
            Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
            A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
            Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
            Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;
            Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
            Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.;
            Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
            M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghm,
            S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
            Guisepi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood,
            C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
            Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
            Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi,
            Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
            Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
            Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
            Maueil, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
            M.; Moesti, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly,
            M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
            V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott,
            A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.;
            Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
            Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon, E.;
            Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
            Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
            B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
            Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
            Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
            Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
            Wambuit, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
            Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
            K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
            Yoshikawa, H.; Danchin, A.
#journal   Nature (1997) 390:249-256
#title     The complete genome sequence of the Gram-positive bacterium
            Bacillus subtilis.
#cross-references MUID:98044033
#accession D69778
#status    preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues  1-219 #label KUN
#cross-references GB:Z99106; GB:AL009126; NID:g2632653; PID:e1182488;
            PID:g2632822
#experimental_source strain 168
GENETICS
#gene      ydeJ
SUMMARY    #length 219 #molecular-weight 24355 #checksum 8514

Query Match 5.9%; Score 6; DB 2; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 143 TOLAGS 148
QY 61 TOLAGS 66

RESULT 34
ENTRY   H64983      #type complete
TITLE   GTP cyclohydrolase I (EC 3.5.4.16) - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE    12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change

```

22-Jun-1999
 H64983; S27052; S18399; A59024
 A64720
#authors
 Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y.
#journal
 Science (1997) 277:1453-1462
#title
 The complete genome sequence of *Escherichia coli* K-12.
#cross-references MUID:97428617
#accession H64983
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-222 #label BLAT
#cross-references GB:AE000304; GB:U00096; NID:g1788470;
#experimental_source PIDN:AACT5214.1; PID:g1788476; UWGP:b2153
#authors Schoedon, G.; Redweik, U.; Frank, G.; Cotton, R.G.H.; Blau, N.
#journal
 Eur. J. Biochem. (1992) 210:561-568
#title
 Allosteric Characteristics of GTP cyclohydrolase I from *Escherichia coli*.
#cross-references MUID:93092993
#accession S27052
#molecule_type protein
#residues 2-46, 'E', 48-51; 100-110, 'X', 112-126, 'X', 128-129 #label SCH1
REFERENCE
#authors S18399
 Katzenmeier, G.; Schmid, C.; Kellermann, J.; Lottspeich, F.; Bacher, A.
#journal
 Biol. Chem. Hoppe-Seyler (1991) 372:991-997
#title
 Biosynthesis of tetrahydrofolate. Sequence of GTP cyclohydrolase I from *Escherichia coli*.
#cross-references MUID:92172284
#accession S18399
#molecule_type DNA
#residues 1-35, 'P', 37-63, 'R', 65-162, 'A', 163-222 #label KAT
#cross-references EMBL:X63910; GB:S85480; NID:g312963
REFERENCE
#authors A59024
 Schmid, C.; Meining, W.; Weinkauff, S.; Bachmann, L.; Ritz, H.; Eberhardt, S.; Gimbel, W.; Werner, T.; Lahm, H.W.; Nar, H.; Bacher, A.
#journal
 Adv. Exp. Med. Biol. (1993) 338:157-162
#title
 Studies on GTP cyclohydrolase I of *Escherichia coli*.
#cross-references MUID:94136208
#accession A59024
#molecule_type DNA
#residues 1-222 #label SCH2
#cross-references GB:X63910; GB:S85480; NID:g312963; PIDN:CAA45365.1; PID:g312964
#experimental_source strain K-12, substrain LE392
#note submitted to EMBL/GenBank/DDBJ by H. Ritz, May 1993
REFERENCE
#authors A65744
 Nar, H.; Huber, R.; Meining, W.; Bacher, A.
#submission submitted to the Brookhaven Protein Data Bank, September 1995
#cross-references PDB:1C9P
#contents annotation; X-ray crystallography, 3.0 angstroms, residues 2-222
REFERENCE
#authors A58904
 Nar, H.; Huber, R.; Auerbach, G.; Fischer, M.; Hoessl, C.; Ritz, H.; Bracher, A.; Meining, W.; Eberhardt, S.; Bacher, A.
#journal
 Proc. Natl. Acad. Sci. U.S.A. (1995) 92:12120-12125
#title
 Active site topology and reaction mechanism of GTP cyclohydrolase I.
#cross-references MUID:96109217
#contents annotation
GENETICS
#gene fole
COMPLEX
FUNCTION homodecamer

#description
 catalyzes the conversion of guanosine triphosphate to dihydroneopterin triphosphate and formate
 folate biosynthesis
CLASSIFICATION
 #superfamily GTP cyclohydrolase I
KEYWORDS
 allosteric regulation; folate biosynthesis; homodecamer; hydrolase
FEATURE
 2-222 #product GTP cyclohydrolase I #status experimental
 111-182 #label MATN
 113,136,180 #disulfide_bonds #status predicted
 #active_site His, Ser, His #status predicted
SUMMARY
 #length 222 #molecular-weight 24830 #checksum 9772
 Query Match 5.9%; Score 6; DB 2; Length 222;
 Best Local Similarity 100.0%; Pred. No. 1.85e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 107 LTSTCE 112
 111111
Qy 18 LTSTCE 23
RESULT 35
ENTRY S29895 #type complete
TITLE GTP cyclohydrolase I (EC 3.5.4.16) - *Escherichia coli*
ORGANISM #formal_name *Escherichia coli*
DATE 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 29-Jan-1999
ACCESSIONS S29895
REFERENCE S29895
 Bacher, A.
#authors submitted to the EMBL Data Library, June 1992
#submission S29895
#accession S29895
#status preliminary
#molecule_type DNA
#residues 1-223 #label BAC
#cross-references EMBL:X63910
CLASSIFICATION #superfamily GTP cyclohydrolase I
KEYWORDS hydrolase
SUMMARY
 #length 223 #molecular-weight 24873 #checksum 3652
 Query Match 5.9%; Score 6; DB 2; Length 223;
 Best Local Similarity 100.0%; Pred. No. 1.85e+01;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 107 LTSTCE 112
 111111
Qy 18 LTSTCE 23
RESULT 36
ENTRY C44355 #type fragment
TITLE CD44 glycoprotein M3 isoform - mouse (fragment)
ORGANISM #formal_name *Mus musculus* #common_name house mouse
DATE 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 07-Feb-1997
ACCESSIONS C44355
REFERENCE A44355
 He, Q.; Lesley, J.; Hyman, R.; Ishihara, K.; Kincade, P.W.
 J. Cell Biol. (1992) 119:1711-1719
#authors Molecular isoforms of murine CD44 and evidence that the membrane proximal domain is not critical for hyaluronate recognition
#journal #title
#cross-references MUID:93107170
#accession C44355
#status preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues 1-224 #label HE1
#experimental_source squamous cell carcinoma KLN205
#note #sequence extracted from NCBI backbone (NCBIP:121205)
KEYWORDS glycoprotein
SUMMARY
 #length 224 #checksum 3958

#authors	Devare, S.G.; Reddy, E.P.; Law, J.D.; Robbins, K.C.; Aaronson, S.A.
#journal	Proc. Natl. Acad. Sci. U.S.A. (1983) 80:731-735
#title	Nucleotide sequence of the simian sarcoma virus genome: demonstration that its acquired cellular sequences encode the transforming gene product p28(sis).
#cross-references	MUID:8314004
#accession	A01381
##molecule_type	genomic RNA
##residues	1-226 ##label DEV
GENETICS	
#gene	sis
CLASSIFICATION	#superfamily platelet-derived growth factor
KEYWORDS	growth factor; transforming protein
FEATURE	
6-226	##domain platelet-derived growth factor chain B similarity #label PDG
SUMMARY	#length 226 #molecular-weight 25411 #checksum 2886
Query Match	5.9%; Score 6; DB 1; Length 226;
Best Local Similarity	100.0%; Pred. NO. 1.85e+01;
Matches	6; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Db	80 AECKTR 85
QY	71 AECKTR 76
RESULT	39
ENTRY	A55030
TITLE	#type fragment platelet-derived growth factor chain B precursor - human (fragment)
ORGANISM	#formal_name Homo sapiens #common_name man
DATE	18-Nov-1994 #sequence_revision 18-Nov-1994 #text_change 17-Mar-1999
ACCESSIONS	A55030
REFERENCE	A55030
#authors	Johnsson, A.; Heidin, C.H.; Wasteson, A.; Westermarck, B.; Deuel, T.F.; Huang, J.S.; Seeburg, P.H.; Gray, A.; Ullrich, A.; Scrace, G.; Stroobant, P.; Waterfield, M.D.
#journal	EMBO J. (1984) 3:921-928
#title	The c-sis gene encodes a precursor of the B chain of platelet-derived growth factor.
#cross-references	MUID:84236121
#accession	A55030
#status	preliminary
##molecule_type	DNA
##residues	1-230 ##label JOH
##cross-references	GB:X00556; GB:X00559; GB:X00560; GB:X00561; GB:X00565
CLASSIFICATION	#superfamily platelet-derived growth factor
SUMMARY	#length 230 #checksum 3580
Query Match	5.9%; Score 6; DB 2; Length 230;
Best Local Similarity	100.0%; Pred. NO. 1.85e+01;
Matches	6; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Db	84 AECKTR 89
QY	71 AECKTR 76
RESULT	40
ENTRY	S11429
TITLE	#type complete nitrate reductase (EC 1.7.99.4) 2 delta chain - Escherichia coli
ALTERNATE_NAMES	narW protein
ORGANISM	#formal_name Escherichia coli
DATE	21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 26-Aug-1999
ACCESSIONS	S11429; E64899
REFERENCE	S11427
#authors	Blasco, F.; Iobbi, C.; Ratouchniak, J.; Bonnefoy, V.; Chipaux, M.

```

#journal Mol. Gen. Genet. (1990) 222:104-111
#title Nitrate reductases of Escherichia coli: sequence of the
        second nitrate reductase and comparison with that encoded
        by the narGHJI operon.
#cross-references MUID:91042410
#accession S11429
##molecule_type DNA
##residues 1-231 ##label BLA
##cross-references EMBL:X17110; NID:g42107; PIDN:CAA34966.1; PID:g42110
REFERENCE A64720
#authors Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
        Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
        Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
        Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
        Y.
#journal Science (1997) 277:1453-1462
#title The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession E64899
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-231 ##label BLAT
##cross-references GB:AE000243; GB:U00096; NID:gl787730;
        PIDN:AC74548.1; PID:gl787739; UMGp:b1466
##experimental_source strain K-12, substrain MG1655
GENETICS
#gene narW
#FUNCTION
#description required for the formation of the active, membrane-bound
        nitrate reductase
CLASSIFICATION #superfamily narJ protein
KEYWORDS nitrate assimilation; oxidoreductase
SUMMARY #length 231 #molecular-weight 26160 #checksum 2808

Query Match 5.9%; Score 6; DB 2; Length 231;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 156 QLAGSS 161
|||||
QY 62 QLAGSS 67

RESULT 41
ENTRY #type complete
TITLE hypothetical protein F41C3_6 - Caenorhabditis elegans
ORGANISM #formal_name Caenorhabditis elegans
DATE 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change
        20-Sep-1999
ACCESSIONS T16315
REFERENCE Chissoe, S.
#authors
#submission submitted to the EMBL Data Library, July 1995
#description The sequence of C. elegans cosmid F41C3.
#accession T16315
##status preliminary; translated from GB/EMBL/DBDJ
##molecule_type DNA
##residues 1-232 ##label CHI
##cross-references EMBL:U23521; NID:g746536; PID:g746542;
        PIDN:AA046813.1; CESP:F41C3.6
##experimental_source strain Bristol N2
GENETICS
#gene CESP:F41C3.6
#introns
SUMMARY #length 232 #molecular-weight 26687 #checksum 6044

Query Match 5.9%; Score 6; DB 2; Length 232;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 146 KFSQTC 151
|||||
QY 3 KFSQTC 8

```

```

RESULT 42
ENTRY #type fragment
TITLE hypothetical protein F10N7.10 - Arabidopsis thaliana
        (fragment)
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
        cress
DATE 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
        21-May-1999
ACCESSIONS T04627
REFERENCE Z15263
#authors Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hoheisel, J.;
        Weves, H.W.; Mayer, K.F.X.; Schueller, C.
#submission submitted to the Protein Sequence Database, March 1999
#accession T04627
##molecule_type DNA
##residues 1-233 ##label BEV
##cross-references EMBL:AL021636
##experimental_source cultivar Columbia; BAC clone F10N7
GENETICS
#map_position 4
#note intron positions not resolved
#note F10N7.10
SUMMARY #length 233 #checksum 5664

Query Match 5.9%; Score 6; DB 2; Length 233;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 211 SSELAA 216
|||||
QY 66 SSELAA 71

RESULT 43
ENTRY #type fragment
TITLE aleurone ribonuclease (EC 3.1.1.-) - barley (fragment)
ORGANISM #formal_name Hordeum vulgare #common_name barley
DATE 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
        22-Jun-1999
ACCESSIONS T04419
REFERENCE Rogers, J.C.; Rogers, S.W.
#authors
#submission submitted to the EMBL Data Library, April 1997
#description Gibberellin-induced barley aleurone RNase.
#accession T04419
##status preliminary; translated from GB/EMBL/DBDJ
##molecule_type mRNA
##residues 1-240 ##label ROG
##cross-references EMBL:AF000939; NID:g2149999; PIDN:AAB58718.1;
        #experimental_source cv. Himalaya
CLASSIFICATION #superfamily Enterobacter ribonuclease
KEYWORDS hydrolase
SUMMARY #length 240 #checksum 3760

Query Match 5.9%; Score 6; DB 2; Length 240;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 65 AECKTR 70
|||||
QY 71 AECKTR 76

RESULT 44
ENTRY #type complete
TITLE platelet-derived growth factor chain B precursor (sis) -
        mouse
ALTERNATE_NAMES PDGF-related transforming protein
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 31-Dec-1992 #sequence_revision 31-Mar-1993 #text_change

```

```

18-Jun-1999
ACCESSIONS A39073
REFERENCE Bonthron, D.T.; Sultan, P.; Collins, T.
#authors Genomics (1991) 10:287-292
#journal Structure of the murine c-sis proto-oncogene (Sis, PDGFB)
#title encoding the B chain of platelet-derived growth factor.
#cross-references MUID:91257844
#accession A39073
##molecule_type DNA
##residues 1-241 ##label BON
##cross-references GB:M64849; GB:M55394; NID:g192818; PIDN:AAA37485.1;
PID:g192820

GENETICS
#gene sis
#superfamily platelet-derived growth factor
#protein glycoprotein; growth factor; platelet; proto-oncogene;
transforming protein
FEATURE
1-20 #domain signal sequence #status predicted #label SIG\
1-81 #domain propeptide #status predicted #label PRO\
82-190 #product platelet-derived growth factor chain B #status
predicted #label MAT\
159-163 #region receptor binding #status predicted\
63 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY
#length 241 #molecular_weight 27381 #checksum 4345
Query Match 5.9%; Score 6; DB 1; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 95 ABEKTR 100
|||||
Qy 71 ABEKTR 76

RESULT 45
ENTRY PFHUG2 #type complete
TITLE platelet-derived growth factor chain B precursor - human
ALTERNATE_NAMES PDGF B-chain; PDGF-B; PDGF-II; PDGF-related transforming
protein (sis)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 18-Apr-1984 #sequence_revision 20-Sep-1984 #text_change
ACCESSIONS A94276; A21024; A23532; A93366; A25141; A94271; A93308;
A43499; S56115; I57635; I37266; A01380; A94622
REFERENCE A94276
#authors Josephs, S.F.; Ratner, L.; Clarke, M.F.; Westin, E.H.; Reitz,
M.S.; Wong-Staal, F.
#journal Science (1984) 225:636-639
#title Transforming potential of human c-sis nucleotide sequences
encoding platelet-derived growth factor.
#cross-references MUID:84250225
#accession A94276
##molecule_type DNA
##residues 1-241 ##label JOS
##cross-references GB:K01401; NID:g338206; PIDN:AAA60552.1; PID:g338209
REFERENCE A21024
#authors Chiu, I.M.; Reddy, E.P.; Givol, D.; Robbins, K.C.; Tronick,
S.R.; Aaronson, S.A.
#journal Cell (1984) 37:123-129
#title Nucleotide sequence analysis identifies the human c-sis
proto-oncogene as a structural gene for platelet-derived
growth factor.
#cross-references MUID:84205633
#accession A21024
##molecule_type DNA
##residues 17-20,'RQ',22-241 ##label CH2
REFERENCE A23532
#authors Rao, C.D.; Igarashi, H.; Chiu, I.M.; Robbins, K.C.; Aaronson,
S.A.
#journal Proc. Natl. Acad. Sci. U.S.A. (1986) 83:2392-2396

#title Structure and sequence of the human c-sis/platelet-derived
growth factor 2 (Sis/PDGF2) transcriptional unit.
#cross-references MUID:86205961
#accession A23532
##molecule_type mRNA
##residues 1-241 ##label RAO
##cross-references GB:M12783; GB:M16288; NID:g338210; PIDN:AAA60553.1;
PID:g338211
REFERENCE A93366
#authors Collins, T.; Ginsburg, D.; Boss, J.M.; Orkin, S.H.; Pober,
J.S.
#journal Nature (1985) 316:748-750
#title Cultured human endothelial cells express platelet-derived
growth factor B chain: cDNA cloning and structural
analysis.
#cross-references MUID:85296313
#accession A93366
##molecule_type mRNA
##residues 1-241 ##label COL
##cross-references GB:X02811; NID:g35371; PIDN:CAA26579.1; PID:g35372
REFERENCE A25141
#authors Weich, H.A.; Sebald, W.; Schairer, H.U.; Hoppe, J.
#journal FEBS Lett. (1986) 198:344-348
#title The human osteosarcoma cell line U-2 OS expresses a 3.8
kilobase mRNA which codes for the sequence of the PDGF-B
chain.
#cross-references MUID:86164981
#accession A25141
##molecule_type mRNA
##residues 26-241 ##label WEI
##cross-references GB:X03702; NID:g35374; PIDN:CAA27333.1; PID:g35375
REFERENCE A94271
#authors Antoniadis, H.N.; Hunkapiller, M.W.
#journal Science (1983) 220:963-965
#title Human platelet-derived growth factor (PDGF): amino-terminal
amino acid sequence.
#cross-references MUID:83197379
#accession A94271
##molecule_type protein
##residues 82-100,'E',102-104,'C',106,'C',108-110 ##label ANT
REFERENCE A93308
#authors Waterfield, M.D.; Scrase, G.T.; Whittle, N.; Stroobant, P.;
Johnsson, A.; Wasteson, A.; Westermark, B.; Heidin, C.H.;
Huang, J.S.; Deuel, T.F.
#journal Nature (1983) 304:35-39
#title Platelet-derived growth factor is structurally related to the
putative transforming protein p28(sis) of simian sarcoma
virus.
#cross-references MUID:83244981
#accession A93308
##molecule_type protein
##residues 82-112 ##label WAT
REFERENCE A43499
#authors Josephs, S.F.; Guo, C.; Ratner, L.; Wong-Staal, F.
#journal Science (1984) 223:487-491
#title Human proto-oncogene nucleotide sequences corresponding to
the transforming region of simian sarcoma virus.
#cross-references MUID:84097555
#accession A43499
#status not compared with conceptual translation
##molecule_type DNA
##residues 'Q',22-241 ##label JO2
REFERENCE S56115
#authors Lu, K.V.; Rohde, M.F.; Thomason, A.R.; Kenney, W.C.; Lu, H.S.
#journal Biochem. J. (1995) 309:411-417
#title Mistranslation of a TGA termination codon as tryptophan in
recombinant platelet-derived growth factor expressed in
Escherichia coli.
#cross-references MUID:95351967
#accession S56115
#status preliminary
##molecule_type protein
##residues 82-93 ##label LUK

```

```

REFERENCE
157635
#authors Rao, C.D.; Pech, M.; Robbins, K.C.; Aaronson, S.A.
#journal Mol. Cell. Biol. (1988) 8:284-292
#title The 5' untranslated sequence of the c-sis/platelet-derived
growth factor 2 transcript is a potent translational
inhibitor.
#cross-references MUID:88094398
#accession 157635
#status translated from GB/EMBL/DBJ
#molecule_type DNA
#residues 1-20 #label RES
#cross-references GB:M19719; NID:g189727; PIDN:AAA60349.1; PID:g553608
REFERENCE
137266
#authors Ratner, L.; Josephs, S.F.; Jarrett, R.; Reitz, M.S.
#journal Nucleic Acids Res. (1985) 13:5007-5018
#title Nucleotide sequence of transforming human c-sis cDNA clones
with homology to platelet-derived growth factor.
#cross-references MUID:85269623
#accession 137266
#status translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-241 #label RE2
#cross-references EMBL:X02744; NID:g30246; PIDN:CAA26524.1; PID:g30247
COMMENT The receptor for this growth factor is a tyrosine kinase.
COMMENT Human platelet-derived growth factor, a potent mitogen for cells of
mesenchymal origin, is a disulfide-linked dimer of two chains,
which may be both of type A, both of type B, or an A-B
heterodimer. Reduction of its disulfide bonds irreversibly
destroys biological activity. Different receptors bind
preferentially to A-A and B-B homodimers. This growth factor
induces a variety of cellular responses.
COMMENT B and A chains are encoded by genes located on different
chromosomes. The two genes are expressed independently in human
tumor cell lines.
GENETICS
#gene GDB:PDGFB
#map_position 22q12.3-22q13.1
#introns 57/3; 94/1; 192/3; 241/1
CLASSIFICATION #superfamily platelet-derived growth factor
KEYWORDS growth factor; mitogen
FEATURE
1-20 #domain signal sequence #status predicted #label SIG\
21-81 #domain amino-terminal propeptide #status predicted
#label PRO\
82-190 #product platelet-derived growth factor chain B #status
experimental #label MAT\
159-163 #region receptor binding #status predicted\
191-241 #domain carboxyl-terminal propeptide #status predicted
#label CTP\
17-141,130-178, #disulfide_bonds #status experimental\
134-180 #disulfide_bonds interchain (to 133 in homodimeric form)
124 #status experimental\
124 #disulfide_bonds interchain (to chain A-132 in
heterodimeric form) #status predicted\
133 #disulfide_bonds interchain (to 124 in homodimeric form)
#status experimental\
133 #disulfide_bonds interchain (to chain A-124 in
heterodimeric form) #status predicted
SUMMARY #length 241 #molecular-weight 27283 #checksum 3348
Query Match 5.9%; Score 6; DB 1; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 95 AECKTR 100
| | | | |
Qy 71 AECKTR 76
RESULT 46 B49276 #type complete
ENTRY

```

```

TITLE 28k protein - Ectromelia virus
ORGANISM #formal_name Ectromelia virus
DATE 01-Dec-1995 #sequence_revision 01-Dec-1995 #text_change
26-Aug-1999
ACCESSIONS B49276
REFERENCE A49276
#authors Senkevich, T.G.; Koonin, E.V.; Buller, R.M.L.
#journal Virology (1994) 198:118-128
#title A poxvirus protein with a RING zinc finger motif is of
crucial importance for virulence.
#cross-references MUID:94082441
#accession B49276
#status preliminary
#molecule_type DNA
#residues 1-241 #label SEN
#cross-references GB:U01161; NID:g397978; PIDN:AAA16258.1; PID:g397980
GENETICS p28
#gene #superfamily vaccinia virus 21.7K HindIII-C protein
CLASSIFICATION #length 241 #molecular-weight 28382 #checksum 5999
SUMMARY
Query Match 5.9%; Score 6; DB 2; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 9 NTSSID 14
| | | | |
Qy 30 NTSSID 35
RESULT 47
ENTRY
TITLE TVCTSS #type complete
ALTERNATE_NAMES platelet-derived growth factor chain B precursor - cat
ORGANISM PDGF-related transforming protein
DATE #formal_name Fells silvestris catu #common_name domestic cat
30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change
31-Mar-1996
ACCESSIONS A26402
REFERENCE A26402
#authors Van den Ouweland, A.M.W.; Van Groningen, J.J.M.; Schalken,
J.A.; Van Neck, H.W.; Bloemers, H.P.J.; Van de Ven, W.J.M.
#journal Nucleic Acids Res. (1987) 15:959-970
#title Genetic organization of the c-sis transcription unit.
#cross-references MUID:87146463
#accession A26402
#molecule_type mRNA
#residues 1-245 #label VAN
GENETICS
#gene sis
CLASSIFICATION #superfamily platelet-derived growth factor
KEYWORDS glycoprotein; growth factor; platelet; proto-oncogene;
transforming protein
FEATURE
1-20 #domain signal sequence #status predicted #label SIG\
21-81 #domain propeptide #status predicted #label PRO\
82-194 #product platelet-derived growth factor chain B #status
predicted #label MAT\
163-167 #region receptor binding #status predicted\
63 #binding_site carbohydrate (Asn) (covalent) #status
predicted
SUMMARY #length 245 #molecular-weight 27787 #checksum 2148
Query Match 5.9%; Score 6; DB 1; Length 245;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 99 AECKTR 104
| | | | |
Qy 71 AECKTR 76
RESULT 48 QQB4L #type complete
ENTRY

```



```

TITLE      probable glycoprotein - human herpesvirus 4 (strain B95-8)
ORGANISM   #formal_name human herpesvirus 4, Epstein-Barr virus
DATE       03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change
          16-Jul-1999
ACCESSIONS A03780; S33052
REFERENCE   A93065
#authors    Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
#journal     Mol. Biol. Med. (1983) 1:21-45
#title      Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8
            Epstein-Barr virus.
#cross-references MUID:85035713
#accession  A03780
#molecule_type DNA
#status     1-248 #label BAN
#residues   ##cross-references EMBL:V01555; NID:g59074; PIDN:CAA24803.1;
            PID:g1334911
REFERENCE   A03794
#authors    Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.;
            Farrell, P.J.; Gibson, T.J.; Hatfull, G.; Hudson, G.S.;
            Satchwell, S.C.; Seguin, C.; Tufnell, P.S.; Barrell, B.G.
#journal     Nature (1984) 310:207-211
#title      DNA sequence and expression of the B95-8 Epstein-Barr virus
            genome.
#cross-references MUID:84270667
#contents   annotation; protein coding region
CLASSIFICATION #superfamily Epstein-Barr virus glycoprotein
KEYWORDS      glycoprotein
FEATURE
27,45,73,83,92,95,
104,116,131,144 #binding_site carbohydrate (Asn) (covalent) #status
SUMMARY      #length 248 #molecular-weight 27076 #checksum 534
            predicted
Query Match      5.9%; Score 6; DB 1; Length 248;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 78 TNGGYN 83
Qy 25 TNGGYN 30
RESULT 49
ENTRY
TITLE      #type complete
ORGANISM   #formal_name Staphylococcus aureus phage phi PVL
DATE       23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change
          23-Apr-1999
ACCESSIONS T00169
REFERENCE   T00169
#authors    Kaneko, J.; Kimura, T.; Kawakami, Y.; Tomita, T.; Kamio, Y.
#journal     Biosci. Biotechnol. Biochem. (1997) 61:1960-1962
#title      Panton-Valentine leukocidin genes in a phage-like particle
            isolated from mitomycin C-treated Staphylococcus aureus V8
            (ATCC 49775).
#accession  T00169
#status     translated from GB/EMBL/DBDJ
#molecule_type DNA
#residues   ##status
#cross-references EMBL:AB009866; NID:d1204727; PID:d1032870
SUMMARY      #length 249 #molecular-weight 28536 #checksum 9098
            5.9%; Score 6; DB 2; Length 249;
Query Match      5.9%; Score 6; DB 2; Length 249;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 60 SVLTST 65
Qy 16 SVLTST 21
T12439 #type complete
RESULT 50
ENTRY
#journal
#title

```

```

TITLE      tonoplast intrinsic protein - common ice plant
ORGANISM   #formal_name Mesembryanthemum crystallinum #common_name
            common ice plant
DATE       23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change
          20-Sep-1999
ACCESSIONS T12439
REFERENCE   Z17519
#authors    Ishitani, M.; Bohnert, H.J.
#submission submitted to the EMBL Data Library, December 1995
#description A transcript for tonoplast water channel protein (TIP) from
            Mesembryanthemum crystallinum.
#accession  T12439
#status     preliminary; translated from GB/EMBL/DBDJ
#molecule_type mRNA
#residues   1-253 #label ISH
#cross-references EMBL:U43291; NID:g1622728; PID:g1622729
GENETICS
#gene       TIP
CLASSIFICATION #superfamily lens fiber membrane major intrinsic protein
KEYWORDS      water channel
SUMMARY      #length 253 #molecular-weight 26080 #checksum 4292
            5.9%; Score 6; DB 2; Length 253;
Query Match      5.9%; Score 6; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 110 QLQSS 115
Qy 62 QLQSS 67
RESULT 51
ENTRY
TITLE      #type complete
ORGANISM   #formal_name Caenorhabditis elegans
DATE       06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
          09-Sep-1997
ACCESSIONS S40702
REFERENCE   S40701
#authors    Smith, M.
#submission submitted to the EMBL Data Library, December 1993
#accession  S40702
#molecule_type DNA
#residues   1-254 #label SMI
#cross-references EMBL:Z29094; NID:g436440; PID:g436442
GENETICS
#introns    61/3; 82/1; 184/3
SUMMARY      #length 254 #molecular-weight 29082 #checksum 3833
            5.9%; Score 6; DB 2; Length 254;
Query Match      5.9%; Score 6; DB 2; Length 254;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 186 LAAECK 191
Qy 69 LAAECK 74
RESULT 52
ENTRY
TITLE      #type complete
ORGANISM   #formal_name Aquifex aeolicus
DATE       08-May-1998 #sequence_revision 08-May-1998 #text_change
          22-Jun-1999
ACCESSIONS B70380
REFERENCE   A70300
#authors    Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.;
            Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.;
            Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short,
            J.W.; Olson, G.J.; Swanson, R.V.
#journal     Nature (1998) 392:353-358
#title      The complete genome of the hyperthermophilic bacterium
            Aquifex aeolicus.

```

```

#cross-references MUID:98196666
#accession B70380
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-255 ##label AOF
#cross-references GB:AE000714; NID:g2983446; PIDN:ARC07032.1;
PID:g2983451; GB:AE000657
##experimental_source strain VF5
GENETICS
#gene rnpH
#superfamily tRNA nucleotidyltransferase
#keywords nucleotidyltransferase; tRNA processing
#length 255 #molecular-weight 28372 #checksum 1544
Query Match 5.9%; Score 6; DB 2; Length 255;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 42 SVIENV 47
|||||
38 SVIENV 43

RESULT 53
ENTRY #type complete
TITLE transcription initiation factor sigma F - Mycobacterium
ALTERNATE_NAMES SigF protein; sporulation sigma factor homolog;
ORGANISM stress-response sigma factor homolog
#formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
18-Jun-1999
ACCESSIONS G70980; JC6131
REFERENCE G70980; JC6131
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eigmeier, K.; Gas, S.; Barry
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from
the complete genome sequence.
#cross-references MUID:98293987
#accession G70980
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-261 ##label COL
#cross-references GB:292771; GB:AL123456; NID:g3242259;
PID:CA807069.1; PID:es06550; PID:g1877341
##experimental_source strain H37Rv
REFERENCE JC6131
#authors Demai, J.; Zhang, Y.; Ko, C.; Young, D.B.; Bishai, W.R.
#journal Proc. Natl. Acad. Sci. U.S.A. (1996) 93:2790-2794
#title A stationary-phase stress-response sigma factor from
Mycobacterium tuberculosis.
#cross-references MUID:96181544
#accession JC6131
#molecule_type DNA
#residues 1-247, 'V', 249-261 ##label DEM
#cross-references GB:U41061; NID:g1276888; PIDN:ARC44103.1;
PID:g1276889
COMMENT This factor is a subunit of bacterial RNA polymerase and confer
promoter specificity to the holoenzyme complex. It plays a role
in the ability of tubercle bacilli to adapt to host defenses and
persist during human infection.
GENETICS
#gene sigF

```

```

#start_codon GTG
CLASSIFICATION #superfamily transcription sigma factor G; transcription
initiation factor sigma katF homolog
KEYWORDS DNA binding; sigma factor; transcription initiation
FEATURE 45-261
SUMMARY #domain transcription initiation factor sigma katF
homology #label KTF
#length 261 #molecular-weight 28793 #checksum 4663
Query Match 5.9%; Score 6; DB 2; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 145 SELAEE 150
|||||
67 SELAEE 72

RESULT 54
ENTRY #type complete
TITLE Mst26Aa protein - fruit fly (Drosophila melanogaster)
ORGANISM #formal_name Drosophila melanogaster
DATE 22-Nov-1993 #sequence_revision 26-May-1995 #text_change
16-Feb-1997
ACCESSIONS S30425
REFERENCE S30407
#authors Aguade, M.; Miyashita, N.; Langley, C.H.
#journal Genetics (1992) 132:755-770
#title Polymorphism and divergence in the Mst26A male accessory
gland gene region in Drosophila.
#cross-references MUID:93106377
#accession S30425
#status preliminary; nucleic acid sequence not shown;
translation not shown
#molecule_type DNA
#residues 1-264 ##label AGU
#cross-references EMBL:X70897
#note the nucleotide sequence was submitted to the EMBL Data
Library, January 1993
GENETICS
#gene FlyBase:Acp26Aa
#introns 12/1
#cross-references FlyBase:FBgn0002855
SUMMARY #length 264 #molecular-weight 29643 #checksum 5378
Query Match 5.9%; Score 6; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 96 SSIDLN 101
|||||
32 SSIDLN 37

RESULT 55
ENTRY #type complete
TITLE male accessory gland secretory protein 26Aa precursor
ALTERNATE_NAMES (variant 1) - fruit fly (Drosophila melanogaster)
ORGANISM male accessory gland secretory protein 355a; Mst26Aa protein
#formal_name Drosophila melanogaster
DATE 02-Dec-1993 #sequence_revision 26-May-1995 #text_change
24-Sep-1998
ACCESSIONS S30413; S30417; S30421
REFERENCE S30407
#authors Aguade, M.; Miyashita, N.; Langley, C.H.
#journal Genetics (1992) 132:755-770
#title Polymorphism and divergence in the Mst26A male accessory
gland gene region in Drosophila.
#cross-references MUID:93106377
#accession S30413
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-264 ##label AGU

```



```

##cross-references EMBL:X70888
##experimental_source allelic variant NCI
REFERENCE S65541
#authors Aguade, M.
#submission submitted to the EMBL Data Library, January 1993
#accession S65541
#molecule_type DNA
#residues_type 1-100,'S',102-264 ##label AGW
##cross-references EMBL:X70888; NID:98232; PID:g8233
GENETICS
#gene FlyBase:Acp26Aa
##cross-references FlyBase:FBgn0002855
#introns 12/1
SUMMARY
#length 264 #molecular-weight 29644 #checksum 5811
Query Match 5.9%; Score 6; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 96 SSIDLN 101
32 SSIDLN 37
RESULT 59
ENTRY A25669 #type complete
TITLE PDGF-related transforming protein (v-sis) - simian sarcoma
ORGANISM #formal_name simian sarcoma virus
DATE 28-Sep-1987 #sequence_revision 28-Sep-1987 #text_change
12-Apr-1995
ACCESSIONS A25669
REFERENCE Hannink, M.; Sauer, M.K.; Donoghue, D.J.
#journal Mol. Cell. Biol. (1986) 6:1304-1314
#cross-references MUID:87064399
#accession A25669
#status preliminary
#molecule_type protein
#residues 1-271 ##label HAN
#note deletions in the C-Terminal Coding Region of the v-sis
Gene: Dimerization Is Required for Transformation
CLASSIFICATION #superfamily platelet-derived growth factor
SUMMARY #length 271 #molecular-weight 30108 #checksum 5973
Query Match 5.9%; Score 6; DB 2; Length 271;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
125 AECKTR 130
71 AECKTR 76
RESULT 60
ENTRY C33282 #type fragment
TITLE DNA-binding protein (clone XLCOF8.4) - African clawed frog
ORGANISM #formal_name Xenopus laevis #common_name African clawed frog
DATE 29-Jan-1990 #sequence_revision 29-Jan-1990 #text_change
06-Dec-1996
ACCESSIONS C33282
REFERENCE Knoechel, W.; Poetting, A.; Koester, M.; El Baradi, T.;
#authors Nietfeld, W.; Bouwmeester, T.; Pieler, T.
#journal Proc. Natl. Acad. Sci. U.S.A. (1989) 86:6097-6100
#title Evolutionary conserved modules associated with zinc fingers
in Xenopus laevis.
#cross-references MUID:89345612
#accession C33282
#status preliminary; nucleic acid sequence not shown; not
compared with conceptual translation
#molecule_type mRNA
##residues 1-284 ##label BAR

##residues 1-273 ##label KNO
KEYWORDS DNA binding; zinc finger
SUMMARY #length 273 #checksum 6523
Query Match 5.9%; Score 6; DB 2; Length 273;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 261 AGSSEL 266
64 AGSSEL 69
RESULT 61
ENTRY A64145 #type complete
TITLE hypothetical protein HI0184 - Haemophilus influenzae (strain
Rd KW20)
ORGANISM #formal_name Haemophilus influenzae
DATE 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
29-Sep-1999
ACCESSIONS A64145
REFERENCE Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.;
Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,
J.L.; Geoghagen, N.S.M.; Gnehm, C.L.; McDonald, L.A.;
Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.C.
Science (1995) 269:496-512
#journal Science
#title Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd.
#cross-references MUID:95350630
#accession A64145
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-275 ##label TIGR
#cross-references GB:U32703; GB:L42023; NID:91573133; PIDN:AAC21853.1;
PID:g1573140; TIGR:HI0184
#note best homolog was a hypothetical protein from Escherichia
coli
CLASSIFICATION #superfamily conserved hypothetical protein YJL068C
SUMMARY #length 275 #molecular-weight 31317 #checksum 805
Query Match 5.9%; Score 6; DB 2; Length 275;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 233 FIETCR 238
54 FIETCR 59
RESULT 62
ENTRY S71851 #type complete
TITLE heat shock transcription factor HSF4 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
cress
DATE 05-Dec-1996 #sequence_revision 13-Mar-1997 #text_change
22-Jun-1999
ACCESSIONS S71851
REFERENCE Barros, D.; Czarnecka-Verner, E.; Yuan, C.X.; Baldwin, D.;
#authors Gurley, W.
#submission submitted to the EMBL Data Library, August 1996
#description Cloning of two cDNAs encoding heat shock transcription
factors from Arabidopsis.
#accession S71851
#molecule_type mRNA
##residues 1-284 ##label BAR

```

Db	125	LTSTCE	130	
QY	18	LTSTCE	23	
RESULT	64			
ENTRY		S44630	#type complete	
TITLE		f22b7.9 protein - Caenorhabditis elegans		
ORGANISM		#formal_name Caenorhabditis elegans		
DATE		20-Feb-1995	#sequence_revision 20-Feb-1995	#text_change

```

Query Match          5.9%;      Score 6;   DB 2;   Length 289;
Best Local Similarity 100.0%;    Pred. No. 1.85e+01;
Matches             6;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;
Db      182 QLAGSS 187
Qy      62 QLAGSS 67
RESULT 66

```

```

ENTRY      E71717      #type complete
TITLE      ribosomal protein S2 - Rickettsia prowazekii
ORGANISM   #formal_name Rickettsia prowazekii
DATE       21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change
26-Aug-1999
ACCESSIONS E71717
REFERENCE   A71630
#authors   Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.;
            Sacheritz-Ponten, T.; Alsmark, U.C.M.; Podowski, R.M.;
            Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.; Kurland,
            C.G.
#journal   Nature (1998) 396:133-140
#title     The genome sequence of Rickettsia prowazekii and the origin
            of mitochondria.
#cross-references MUID:99039499
#accession E71717
#status    preliminary; nucleic acid sequence not shown;
            translation not shown
#molecule_type DNA
#residues  1-296 #label AND
#cross-references GB:AJ235270; GB:AJ235269; NID:g3860572; PID:e1342399;
            PID:g3860655
#experimental_source strain Madrid E
GENETICS
#gene      rpsB; RP086
CLASSIFICATION
#length 296 #molecular-weight 32961 #checksum 4126
SUMMARY
Query Match      5.9%; Score 6; DB 2; Length 296;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 72 FVSTKI 77
      |||||
Qy 80 FVSTKI 85

RESULT      67
ENTRY
TITLE      D44355      #type fragment
ORGANISM   CD44 glycoprotein M4 isoform - mouse (fragment)
DATE       30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change
07-Feb-1997
ACCESSIONS D44355
REFERENCE   A44355
#authors   He, Q.; Lesley, J.; Hyman, R.; Ishihara, K.; Kincade, P.W.
#journal   J. Cell Biol. (1992) 119:1711-1719
#title     Molecular isoforms of murine CD44 and evidence that the
            membrane proximal domain is not critical for hyaluronate
            recognition.
#cross-references MUID:93107170
#accession D44355
#status    preliminary; not compared with conceptual translation
#molecule_type mRNA
#residues  1-301 #label HE1
#experimental_source squamous cell carcinoma KLN205
#note      sequence extracted from NCBI backbone (NCBIP:121207)
KEYWORDS    glycoprotein
SUMMARY     #length 301 #checksum 889
Query Match      5.9%; Score 6; DB 2; Length 301;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 290 NVDSGL 295
      |||||
Qy 42 NVDSGL 47

RESULT      68
ENTRY
TITLE      T02638      #type complete
            G2 protein homolog - slime mold (Dictyostelium discoideum)
            plasmid bdp5

```

```

ORGANISM   #formal_name Dictyostelium discoideum
DATE       24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change
07-May-1999
ACCESSIONS T02638
REFERENCE   Z14884
#authors   Rieben Jr., W.K.; Gonzales, C.M.; Gonzales, S.T.; Pilkington,
            K.J.; Kiyosawa, H.; Hughes, J.E.; Welker, D.L.
#journal   Genetics (1998) 148:1117-1125
#title     Dictyostelium discoideum nuclear plasmid Ddp5 is a chimera
            related to the Ddpl and Ddp2 plasmid families.
#cross-references MUID:98198836
#accession T02638
#status    preliminary; translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues  1-311 #label R1E
#cross-references EMBL:AF000580; NID:g3068582; PID:g3068587
#experimental_source strain WS2162
GENETICS
#gene      q2
#plasmid   Ddp5
SUMMARY     #length 311 #molecular-weight 36889 #checksum 8773
Query Match      5.9%; Score 6; DB 2; Length 311;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 110 VSTKIN 115
      |||||
Qy 81 VSTKIN 86

RESULT      69
ENTRY
TITLE      S57546      #type complete
            hypothetical protein YPR013c - yeast (Saccharomyces
            cerevisiae)
ALTERNATE_NAMES
ORGANISM   hypothetical protein LP212c; hypothetical protein YP9531.06c
DATE       10-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
12-Dec-1997
ACCESSIONS S57546; S59757
REFERENCE   S57541
#authors   Bowman, S.
#submission submitted to the EMBL Data Library, June 1995
#accession S57546
#molecule_type DNA
#residues  1-317 #label BOW
#cross-references EMBL:Z49919; NID:g887584; PID:g887590; MIPS:YPR013c
#experimental_source strain AB972
REFERENCE   S59746
#authors   Wang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.;
            Hall, J.; Storms, R.K.; Vo, D.H.; Winnett, E.
#submission submitted to the EMBL Data Library, July 1995
#description The sequence of Saccharomyces cerevisiae chromosome XVI right
            arm.
#accession S59757
#molecule_type DNA
#residues  1-317 #label WAN
#cross-references EMBL:U31900; NID:g1276597; PID:g939746; MIPS:YPR013c
GENETICS
SUMMARY     #map_position 16R
            #length 317 #molecular-weight 35358 #checksum 8948
Query Match      5.9%; Score 6; DB 2; Length 317;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 69 STKINL 74
      |||||
Qy 82 STKINL 87

RESULT      70
ENTRY
TITLE      S54361      #type complete

```

```

TITLE      gag-related protein - Trypanosoma cruzi
ORGANISM   #formal_name Trypanosoma cruzi
DATE       15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change
REFERENCE  07-May-1999
#authors   S54361
#journal   S54359
#title     Martin, F.; Maranon, C.; Olivares, M.; Alonso, C.; Lopez,
           M.C.
           J. Mol. Biol. (1995) 247:49-59
           Characterization of a non-long terminal repeat
           retrotransposon cDNA (LITc) from Trypanosoma cruzi:
           homology of the first ORF with the Ape family of DNA repair
           enzymes.
#cross-references MUID:95205412
#accession  S54361
#status    preliminary
#molecule_type mRNA
#residues  1-324 #label MAR
#cross-references EMBL:X83098
PICS
start_codon GTG
SUMMARY    #length 324 #molecular-weight 36973 #checksum 5843

Query Match      5.9%; Score 6; DB 2; Length 324;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 294 LGKFSQ 299
QY 1 LGKFSQ 6

RESULT 71
ENTRY   D64785 #type complete
TITLE   sfmH protein precursor - Escherichia coli
ORGANISM #formal_name Escherichia coli
DATE    12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
17-Jul-1998
ACCESSIONS D64785
REFERENCE  A64720
#authors   Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
           Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
           Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
           Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
           Y.
           Science (1997) 277:1453-1462
           The complete genome sequence of Escherichia coli K-12.
#journal   Science
#title     The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession D64785
#status    nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues  1-325 #label BLAT
#cross-references GB:AE000159; GB:U00096; NID:g1786739; PID:g1786745;
           UWGP:B0533
#experimental_source strain K-12, substrain MG1655

GENETICS
#gene      sfmH
SUMMARY    #length 325 #molecular-weight 35493 #checksum 2967

Query Match      5.9%; Score 6; DB 2; Length 325;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 278 DLNSVI 283
QY 35 DLNSVI 40

RESULT 72
ENTRY   JS0459 #type complete
TITLE   gene H protein - phage SL3
ORGANISM #formal_name phage SL3
#note    host Escherichia coli

```

```

DATE       31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change
ACCESSIONS JS0459
REFERENCE  JS0450
#authors   Lau, P.C.K.; Spencer, J.H.
#journal   Gene (1985) 40:273-284
#title     Nucleotide sequence and genome organization of bacteriophage
           SL3 DNA.
#cross-references MUID:86165869
#accession  JS0459
#molecule_type DNA
#residues  1-328 #label LAU
#cross-references EMBL:M14428; NID:g216089; PIDN:AAA32592.1;
           PID:g216101
CLASSIFICATION #superfamily phage phi-X174 gene H protein
SUMMARY        #length 328 #molecular-weight 34405 #checksum 4113

Query Match      5.9%; Score 6; DB 2; Length 328;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 54 SAIQGS 59
QY 11 SAIQGS 16

RESULT 73
ENTRY   ZHBPP4 #type complete
TITLE   gene H protein - phage phi-X174
ORGANISM #formal_name phage phi-X174
DATE    30-Apr-1980 #sequence_revision 30-Apr-1980 #text_change
23-Jul-1999
ACCESSIONS D93185; A04254
REFERENCE  A93185
#authors   Sanger, F.; Air, G.M.; Barrell, B.G.; Brown, N.L.; Coulson,
           A.R.; Fiddes, J.C.; Hutchison III, C.A.; Slocombe, P.M.;
           Smith, M.
           Nature (1977) 265:687-695
           Nucleotide sequence of bacteriophage phiX174 DNA.
#journal   Nature
#title     Nucleotide sequence of bacteriophage phiX174 DNA.
#cross-references MUID:77171175
#accession D93185
#molecule_type DNA
#residues  1-328 #label SAN
#cross-references GB:J02482; GB:M10348; GB:M10379; GB:M10714;
           GB:M10749; GB:M10750; GB:M10866; GB:M10867;
           GB:M24859; NID:g216019; PIDN:AAA32580.1; PID:g216030

REFERENCE  A94690
#authors   Sanger, F.; Coulson, A.R.; Friedmann, T.; Air, G.M.; Barrell,
           B.G.; Brown, N.L.; Fiddes, J.C.; Hutchison III, C.A.;
           Slocombe, P.M.; Smith, M.
           J. Mol. Biol. (1978) 125:225-246
           The nucleotide sequence of bacteriophage phiX174.
#journal   J. Mol. Biol.
#title     The nucleotide sequence of bacteriophage phiX174.
#cross-references MUID:79091185
#contents  annotation; 30 revisions to the nucleotide sequence of the
           genome
COMMENT     Gene H protein is the minor spike component of the viral shell.
CLASSIFICATION #superfamily phage phi-X174 gene H protein
SUMMARY        #length 328 #molecular-weight 34419 #checksum 3614

Query Match      5.9%; Score 6; DB 1; Length 328;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 54 SAIQGS 59
QY 11 SAIQGS 16

RESULT 74
ENTRY   ZHBPA3 #type complete
TITLE   gene H protein - phage alpha-3
ORGANISM #formal_name phage alpha-3
DATE    17-Dec-1982 #sequence_revision 19-Apr-1996 #text_change

```



```

ACCESSIONS
REFERENCE
#authors
#journal
#title
#cross-references MUID:85226468
#accession A21537
#status preliminary
#molecule_type DNA
#residues_type 1-330 #label KOD
#cross-references GB:M25640; NID:g166101; PIDN:AAA32174.1; PID:g166102
S22324
#authors Kodaira, K.I.; Nakano, K.; Okada, S.; Taketo, A.
#journal Biochim. Biophys. Acta (1992) 1130:277-288
#title Nucleotide sequence of the genome of the bacteriophage
alpha3: interrelationship of the genome structure and the
gene products with those of the phages, phiX174, G4 and
phiK.
#cross-references MUID:92223109
#accession S22332
#molecule_type DNA
#residues_type 1-330 #label KOD
#cross-references EMBL:X60322; NID:g14775; PIDN:CAA42883.1; PID:g14785
A92247
#authors Sims, J.; Capon, D.; Dressler, D.
#journal J. Biol. Chem. (1979) 254:12615-12628
#title dnaG (primase)-dependent origins of DNA replication.
Nucleotide sequences of the negative strand initiation
sites of bacteriophages St-1, phiK, and alpha3.
#cross-references MUID:80049950
#accession A04257
#molecule_type DNA
#residues_type 1-70 #label SIM
#cross-references GB:J02444; GB:M10725; NID:g166103; PIDN:AAA32176.1;
PID:g166105

GENETICS
#gene H
#FUNCTION gene H protein is a minor spike component of the viral shell
CLASSIFICATION #superfamily phage phi-X174 gene H protein
SUMMARY #length 330 #molecular-weight 34844 #checksum 6269

Query Match 5.9%; Score 6; DB 1; Length 330;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

P 58 SAIQGS 63
Q 11 SAIQGS 16

RESULT 75
ENTRY JC4807 #type complete
TITLE core protein H - phage phi-K
ORGANISM #formal_name phage phi-K
DATE 15-Aug-1996 #sequence_revision 15-Oct-1996 #text_change
20-Sep-1999
ACCESSIONS JC4807; B04256; A04256
REFERENCE JC4804
#authors Kodaira, K.; Oki, M.; Kakikawa, M.; Kimoto, H.; Taketo, A.
#journal J. Biochem. (1996) 119:1062-1069
#title The virion proteins encoded by bacteriophage phi-K and its
host-range mutant phi-Khm: Host-range determination and DNA
binding properties.
#cross-references MUID:96424987
#accession JC4807
#status nucleic acid sequence not shown
#molecule_type DNA
#residues 1-332 #label KOD
#cross-references EMBL:X60323; NID:g1478118; PIDN:CAA42893.1;
PID:ei24678; PID:g1478128

```

```

A92247
#authors Sims, J.; Capon, D.; Dressler, D.
#journal J. Biol. Chem. (1979) 254:12615-12628
#title dnaG (primase)-dependent origins of DNA replication.
Nucleotide sequences of the negative strand initiation
sites of bacteriophages St-1, phiK, and alpha3.
#cross-references MUID:80049950
#accession B04256
#molecule_type DNA
#residues_type 1-42 #label SIM
#cross-references GB:M10726; NID:g215500; PIDN:AAA32365.1; PID:g553016
COMMENT This protein is a minor spike component of the viral shell.
GENETICS
#gene H
#CLASSIFICATION #superfamily phage phi-X174 gene H protein
#KEYWORDS spike protein
#SUMMARY #length 332 #molecular-weight 35110 #checksum 229

Query Match 5.9%; Score 6; DB 2; Length 332;
Best Local Similarity 100.0%; Pred. No. 1.85e+01;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 60 SAIQGS 65
Q 11 SAIQGS 16

Search completed: Wed Aug 16 09:58:07 2000
Job time : 48 secs.

```